(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 August 2002 (08.08.2002)

PCT

(10) International Publication Number
WO 02/060942 A2

(51) International Patent Classification[7]: C07K 14/435

(21) International Application Number: PCT/US02/02813

(22) International Filing Date: 30 January 2002 (30.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/265,705	31 January 2001 (31.01.2001)	US
60/266,762	5 February 2001 (05.02.2001)	US
60/269,581	16 February 2001 (16.02.2001)	US
60/271,198	23 February 2001 (23.02.2001)	US
60/272,813	1 March 2001 (01.03.2001)	US
60/275,586	13 March 2001 (13.03.2001)	US
60/278,505	23 March 2001 (23.03.2001)	US
60/280,539	30 March 2001 (30.03.2001)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WARREN, Bridget, A.** [US/US]; 10130 Parkwood Drive, #2, Cupertino, CA 95014 (US). **HONCHELL, Cynthia, D.** [US/US]; 400 Laurel Street #203, San Carlos, CA 94070 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street, #205, San Leandro, CA 94577 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **DELEGEANE, Angelo, M.** [US/US]; 594 Angus Drive, Milpitas, CA 95035 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **TANG, Tom, Y.** [US/US]; 4230 Ramwick Court, San Jose, CA 95118 (US). **LAL, Preeti., G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **BOROWSKI, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **DUGGAN, Brendan, M.** [US/US]; 243 Buena Vista Avenue, #306, Sunnyvale, CA 94086 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **ARVIZU, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **DING, Li** [CN/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **LEE, Sally** [US/US]; 825 East Evelyn #425, Sunnyvale, CA 94086 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES

(57) Abstract: The invention provides human protein modification and maintenance molecules (PMMM) and polynucleotides which identify and encode PMMM. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PMMM.

WO 02/060942 A2

PROTEIN MODIFICATION AND MAINTENANCE MOLECULES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of protein modification and maintenance molecules and to the use of these sequences in the diagnosis, treatment, and prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of protein modification and maintenance molecules.

BACKGROUND OF THE INVENTION

Proteases cleave proteins and peptides at the peptide bond that forms the backbone of the protein or peptide chain. Proteolysis is one of the most important and frequent enzymatic reactions that occurs both within and outside of cells. Proteolysis is responsible for the activation and maturation of nascent polypeptides, the degradation of misfolded and damaged proteins, and the controlled turnover of peptides within the cell. Proteases participate in digestion, endocrine function, and tissue remodeling during embryonic development, wound healing, and normal growth. Proteases can play a role in regulatory processes by affecting the half life of regulatory proteins. Proteases are involved in the etiology or progression of disease states such as inflammation, angiogenesis, tumor dispersion and metastasis, cardiovascular disease, neurological disease, and bacterial, parasitic, and viral infections.

Proteases can be categorized on the basis of where they cleave their substrates. Exopeptidases, which include aminopeptidases, dipeptidyl peptidases, tripeptidases, carboxypeptidases, peptidyl-di-peptidases, dipeptidases, and omega peptidases, cleave residues at the termini of their substrates. Endopeptidases, including serine proteases, cysteine proteases, and metalloproteases, cleave at residues within the peptide. Four principal categories of mammalian proteases have been identified based on active site structure, mechanism of action, and overall three-dimensional structure. (See Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 1-5.)

Serine Proteases

The serine proteases (SPs) are a large, widespread family of proteolytic enzymes that include the digestive enzymes trypsin and chymotrypsin, components of the complement and blood-clotting cascades, and enzymes that control the degradation and turnover of macromolecules within the cell and in the extracellular matrix. Most of the more than 20 subfamilies can be grouped into six clans, each with a common ancestor. These six clans are hypothesized to have descended from at least four

evolutionarily distinct ancestors. SPs are named for the presence of a serine residue found in the active catalytic site of most families. The active site is defined by the catalytic triad, a set of conserved asparagine, histidine, and serine residues critical for catalysis. These residues form a charge relay network that facilitates substrate binding. Other residues outside the active site form an oxyanion hole that stabilizes the tetrahedral transition intermediate formed during catalysis. SPs have a wide range of substrates and can be subdivided into subfamilies on the basis of their substrate specificity. The main subfamilies are named for the residue(s) after which they cleave: trypases (after arginine or lysine), aspases (after aspartate), chymases (after phenylalanine or leucine), metases (methionine), and serases (after serine) (Rawlings, N.D. and A.J. Barrett (1994) Meth. Enzymol. 244:19-61).

Most mammalian serine proteases are synthesized as zymogens, inactive precursors that are activated by proteolysis. For example, trypsinogen is converted to its active form, trypsin, by enteropeptidase. Enteropeptidase is an intestinal protease that removes an N-terminal fragment from trypsinogen. The remaining active fragment is trypsin, which in turn activates the precursors of the other pancreatic enzymes. Likewise, proteolysis of prothrombin, the precursor of thrombin, generates three separate polypeptide fragments. The N-terminal fragment is released while the other two fragments, which comprise active thrombin, remain associated through disulfide bonds.

The two largest SP subfamilies are the chymotrypsin (S1) and subtilisin (S8) families. Some members of the chymotrypsin family contain two structural domains unique to this family. Kringle domains are triple-looped, disulfide cross-linked domains found in varying copy number. Kringles are thought to play a role in binding mediators such as membranes, other proteins or phospholipids, and in the regulation of proteolytic activity (PROSITE PDOC00020). Apple domains are 90 amino-acid repeated domains, each containing six conserved cysteines. Three disulfide bonds link the first and sixth, second and fifth, and third and fourth cysteines (PROSITE PDOC00376). Apple domains are involved in protein-protein interactions. S1 family members include trypsin, chymotrypsin, coagulation factors IX-XII, complement factors B, C, and D, granzymes, kallikrein, and tissue- and urokinase-plasminogen activators. The subtilisin family has members found in the eubacteria, archaebacteria, eukaryotes, and viruses. Subtilisins include the proprotein-processing endopeptidases kexin and furin and the pituitary prohormone convertases PC1, PC2, PC3, PC6, and PACE4 (Rawlings and Barrett, supra).

SPs have functions in many normal processes and some have been implicated in the etiology or treatment of disease. Enterokinase, the initiator of intestinal digestion, is found in the intestinal brush border, where it cleaves the acidic propeptide from trypsinogen to yield active trypsin (Kitamoto, Y. et al. (1994) Proc. Natl. Acad. Sci. USA 91:7588-7592). Prolylcarboxypeptidase, a lysosomal serine peptidase that cleaves peptides such as angiotensin II and III and [des-Arg9]

bradykinin, shares sequence homology with members of both the serine carboxypeptidase and prolylendopeptidase families (Tan, F. et al. (1993) J. Biol. Chem. 268:16631-16638). The protease neuropsin may influence synapse formation and neuronal connectivity in the hippocampus in response to neural signaling (Chen, Z.-L. et al. (1995) J Neurosci 15:5088-5097). Tissue plasminogen activator is useful for acute management of stroke (Zivin, J.A. (1999) Neurology 53:14-19) and myocardial infarction (Ross, A.M. (1999) Clin. Cardiol. 22:165-171). Some receptors (PAR, for proteinase-activated receptor), highly expressed throughout the digestive tract, are activated by proteolytic cleavage of an extracellular domain. The major agonists for PARs, thrombin, trypsin, and mast cell tryptase, are released in allergy and inflammatory conditions. Control of PAR activation by proteases has been suggested as a promising therapeutic target (Vergnolle, N. (2000) Aliment. Pharmacol. Ther. 14:257-266; Rice, K.D. et al. (1998) Curr. Pharm. Des. 4:381-396). Tryptases, the predominant proteins of human mast cells, have been implicated as pathogenetic mediators of allergic and inflammatory conditions, most notably asthma. Properties that distinguish tryptases among the serine proteinases include their activity as heparin-stabilized tetramers, their resistance to many proteinaceous inhibitors, and their preference for peptidergic over macromolecular substrates (Sommerhoff, C.P. et al. (2000) Biochim. Biophys. Acta 1477:75-89).

Prostate-specific antigen (PSA) is a kallikrein-like serine protease synthesized and secreted exclusively by epithelial cells in the prostate gland. Serum PSA is elevated in prostate cancer and is the most sensitive physiological marker for monitoring cancer progression and response to therapy. PSA can also identify the prostate as the origin of a metastatic tumor (Brawer, M.K. and P.H. Lange (1989) Urology 33:11-16).

The signal peptidase is a specialized class of SP found in all prokaryotic and eukaryotic cell types that serves in the processing of signal peptides from certain proteins. Signal peptides are amino-terminal domains of a protein which direct the protein from its ribosomal assembly site to a particular cellular or extracellular location. Once the protein has been exported, removal of the signal sequence by a signal peptidase and posttranslational processing, e.g., glycosylation or phosphorylation, activate the protein. Signal peptidases exist as multi-subunit complexes in both yeast and mammals. The canine signal peptidase complex is composed of five subunits, all associated with the microsomal membrane and containing hydrophobic regions that span the membrane one or more times (Shelness, G.S. and G. Blobel (1990) J. Biol. Chem. 265:9512-9519). Some of these subunits serve to fix the complex in its proper position on the membrane while others contain the actual catalytic activity.

Another family of proteases which have a serine in their active site are dependent on the hydrolysis of ATP for their activity. These proteases contain proteolytic core domains and regulatory ATPase domains which can be identified by the presence of the P-loop, an ATP/GTP-binding motif

(PROSITE PDOC00803). Members of this family include the eukaryotic mitochondrial matrix proteases, Clp protease and the proteasome. Clp protease was originally found in plant chloroplasts but is believed to be widespread in both prokaryotic and eukaryotic cells. The gene for early-onset torsion dystonia encodes a protein related to Clp protease (Ozelius, L.J. et al. (1998) Adv. Neurol. 78:93-105).

The proteasome is an intracellular protease complex found in some bacteria and in all eukaryotic cells, and plays an important role in cellular physiology. Proteasomes are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins of all types, including proteins that function to activate or repress cellular processes such as transcription and cell cycle progression (Ciechanover, A. (1994) Cell 79:13-21). In the UCS pathway, proteins targeted for degradation are conjugated to ubiquitin, a small heat stable protein. The ubiquitinated protein is then recognized and degraded by the proteasome. The resultant ubiquitin-peptide complex is hydrolyzed by a ubiquitin carboxyl terminal hydrolase, and free ubiquitin is released for reutilization by the UCS. Ubiquitin-proteasome systems are implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes (p53), cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, supra). This pathway has been implicated in a number of diseases, including cystic fibrosis, Angelman's syndrome, and Liddle syndrome (reviewed in Schwartz, A.L. and A. Ciechanover (1999) Annu. Rev. Med. 50:57-74). A murine proto-oncogene, Unp, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH3T3 cells. The human homologue of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) Oncogene 10:2179-2183). Ubiquitin carboxyl terminal hydrolase is involved in the differentiation of a lymphoblastic leukemia cell line to a non-dividing mature state (Maki, A. et al. (1996) Differentiation 60:59-66). In neurons, ubiquitin carboxyl terminal hydrolase (PGP 9.5) expression is strong in the abnormal structures that occur in human neurodegenerative diseases (Lowe, J. et al. (1990) J. Pathol. 161:153-160). The proteasome is a large (~2000 kDa) multisubunit complex composed of a central catalytic core containing a variety of proteases arranged in four seven-membered rings with the active sites facing inwards into the central cavity, and terminal ATPase subunits covering the outer port of the cavity and regulating substrate entry (for review, see Schmidt, M. et al. (1999) Curr. Opin. Chem. Biol. 3:584-591).

Cysteine Proteases

Cysteine proteases (CPs) are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Nearly half of the CPs known are present only in viruses. CPs have a cysteine as the major catalytic residue at the active site where catalysis proceeds via a thioester intermediate and is facilitated by nearby histidine and asparagine

residues. A glutamine residue is also important, as it helps to form an oxyanion hole. Two important CP families include the papain-like enzymes (C1) and the calpains (C2). Papain-like family members are generally lysosomal or secreted and therefore are synthesized with signal peptides as well as propeptides. Most members bear a conserved motif in the propeptide that may have structural significance (Karrer, K.M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:3063-3067). Three-dimensional structures of papain family members show a bilobed molecule with the catalytic site located between the two lobes. Papains include cathepsins B, C, H, L, and S, certain plant allergens and dipeptidyl peptidase (for a review, see Rawlings, N.D. and A.J. Barrett (1994) Meth. Enzymol. 244:461-486).

Some CPs are expressed ubiquitously, while others are produced only by cells of the immune system. Of particular note, CPs are produced by monocytes, macrophages and other cells which migrate to sites of inflammation and secrete molecules involved in tissue repair. Overabundance of these repair molecules plays a role in certain disorders. In autoimmune diseases such as rheumatoid arthritis, secretion of the cysteine peptidase cathepsin C degrades collagen, laminin, elastin and other structural proteins found in the extracellular matrix of bones. Bone weakened by such degradation is also more susceptible to tumor invasion and metastasis. Cathepsin L expression may also contribute to the influx of mononuclear cells which exacerbates the destruction of the rheumatoid synovium (Keyszer, G.M. (1995) Arthritis Rheum. 38:976-984).

Calpains are calcium-dependent cytosolic endopeptidases which contain both an N-terminal catalytic domain and a C-terminal calcium-binding domain. Calpain is expressed as a proenzyme heterodimer consisting of a catalytic subunit unique to each isoform and a regulatory subunit common to different isoforms. Each subunit bears a calcium-binding EF-hand domain. The regulatory subunit also contains a hydrophobic glycine-rich domain that allows the enzyme to associate with cell membranes. Calpains are activated by increased intracellular calcium concentration, which induces a change in conformation and limited autolysis. The resultant active molecule requires a lower calcium concentration for its activity (Chan, S.L. and M.P. Mattson (1999) J. Neurosci. Res. 58:167-190). Calpain expression is predominantly neuronal, although it is present in other tissues. Several chronic neurodegenerative disorders, including ALS, Parkinson's disease and Alzheimer's disease are associated with increased calpain expression (Chan and Mattson, supra). Calpain-mediated breakdown of the cytoskeleton has been proposed to contribute to brain damage resulting from head injury (McCracken, E. et al. (1999) J. Neurotrauma 16:749-761). Calpain-3 is predominantly expressed in skeletal muscle, and is responsible for limb-girdle muscular dystrophy type 2A (Minami, N. et al. (1999) J. Neurol. Sci. 171:31-37).

Another family of thiol proteases is the caspases, which are involved in the initiation and execution phases of apoptosis. A pro-apoptotic signal can activate initiator caspases that trigger a

proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death of the cell. Two active site residues, a cysteine and a histidine, have been implicated in the catalytic mechanism. Caspases are among the most specific endopeptidases, cleaving after aspartate residues. Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain interacts with cofactors that can positively or negatively affect apoptosis. An activating signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1 numbering convention) and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long prodomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. Some caspases contain a "death effector domain" in their prodomain by which they can be recruited into self-activating complexes with other caspases and FADD protein associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer. Endogenous caspase inhibitors (inhibitor of apoptosis proteins, or IAPs) also exist. All these interactions have clear effects on the control of apoptosis (reviewed in Chan and Mattson, supra; Salveson, G.S. and V.M. Dixit (1999) Proc. Natl. Acad. Sci. USA 96:10964-10967).

Caspases have been implicated in a number of diseases. Mice lacking some caspases have severe nervous system defects due to failed apoptosis in the neuroepithelium and suffer early lethality. Others show severe defects in the inflammatory response, as caspases are responsible for processing IL-1b and possibly other inflammatory cytokines (Chan and Mattson, supra). Cowpox virus and baculoviruses target caspases to avoid the death of their host cell and promote successful infection. In addition, increases in inappropriate apoptosis have been reported in AIDS, neurodegenerative diseases and ischemic injury, while a decrease in cell death is associated with cancer (Salveson and Dixit, supra; Thompson, C.B. (1995) Science 267:1456-1462).

Aspartyl proteases

Aspartyl proteases (APs) include the lysosomal proteases cathepsins D and E, as well as chymosin, renin, and the gastric pepsins. Most retroviruses encode an AP, usually as part of the pol polyprotein. APs, also called acid proteases, are monomeric enzymes consisting of two domains, each domain containing one half of the active site with its own catalytic aspartic acid residue. APs are most active in the range of pH 2–3, at which one of the aspartate residues is ionized and the other neutral. The pepsin family of APs contains many secreted enzymes, and all are likely to be synthesized with signal peptides and propeptides. Most family members have three disulfide loops, the first ~5 residue loop following the first aspartate, the second 5-6 residue loop preceding the second aspartate, and the third and largest loop occurring toward the C terminus. Retropepsins, on

the other hand, are analogous to a single domain of pepsin, and become active as homodimers with each retropepsin monomer contributing one half of the active site. Retropepsins are required for processing the viral polyproteins.

APs have roles in various tissues, and some have been associated with disease. Renin mediates the first step in processing the hormone angiotensin, which is responsible for regulating electrolyte balance and blood pressure (reviewed in Crews, D.E. and S.R. Williams (1999) Hum. Biol. 71:475-503). Abnormal regulation and expression of cathepsins are evident in various inflammatory disease states. Expression of cathepsin D is elevated in synovial tissues from patients with rheumatoid arthritis and osteoarthritis. The increased expression and differential regulation of the cathepsins are linked to the metastatic potential of a variety of cancers (Chambers, A.F. et al. (1993) Crit. Rev. Oncol. 4:95-114).

Metalloproteases

Metalloproteases require a metal ion for activity, usually manganese or zinc. Examples of manganese metalloenzymes include aminopeptidase P and human proline dipeptidase (PEPD). Aminopeptidase P can degrade bradykinin, a nonapeptide activated in a variety of inflammatory responses. Aminopeptidase P has been implicated in coronary ischemia/reperfusion injury. Administration of aminopeptidase P inhibitors has been shown to have a cardioprotective effect in rats (Ersahin, C. et al (1999) J. Cardiovasc. Pharmacol. 34:604-611).

Most zinc-dependent metalloproteases share a common sequence in the zinc-binding domain. The active site is made up of two histidines which act as zinc ligands and a catalytic glutamic acid C-terminal to the first histidine. Proteins containing this signature sequence are known as the metzincins and include aminopeptidase N, angiotensin-converting enzyme, neurolysin, the matrix metalloproteases and the adamalysins (ADAMS). An alternate sequence is found in the zinc carboxypeptidases, in which all three conserved residues – two histidines and a glutamic acid – are involved in zinc binding.

A number of the neutral metalloendopeptidases, including angiotensin converting enzyme and the aminopeptidases, are involved in the metabolism of peptide hormones. High aminopeptidase B activity, for example, is found in the adrenal glands and neurohypophyses of hypertensive rats (Prieto, I. et al. (1998) Horm. Metab. Res. 30:246-248). Oligopeptidase M/neurolysin can hydrolyze bradykinin as well as neurotensin (Serizawa, A. et al. (1995) J. Biol. Chem 270:2092-2098). Neurotensin is a vasoactive peptide that can act as a neurotransmitter in the brain, where it has been implicated in limiting food intake (Tritos, N.A. et al. (1999) Neuropeptides 33:339-349).

The matrix metalloproteases (MMPs) are a family of at least 23 enzymes that can degrade components of the extracellular matrix (ECM). They are Zn^{+2} endopeptidases with an N-terminal catalytic domain. Nearly all members of the family have a hinge peptide and C-terminal domain

which can bind to substrate molecules in the ECM or to inhibitors produced by the tissue (TIMPs, for tissue inhibitor of metalloprotease; Campbell, I.L. et al. (1999) Trends Neurosci. 22:285). The presence of fibronectin-like repeats, transmembrane domains, or C-terminal hemopexinase-like domains can be used to separate MMPs into collagenase, gelatinase, stromelysin and membrane-type MMP subfamilies. In the inactive form, the Zn^{+2} ion in the active site interacts with a cysteine in the pro-sequence. Activating factors disrupt the Zn^{+2}-cysteine interaction, or "cysteine switch," exposing the active site. This partially activates the enzyme, which then cleaves off its propeptide and becomes fully active. MMPs are often activated by the serine proteases plasmin and furin. MMPs are often regulated by stoichiometric, noncovalent interactions with inhibitors; the balance of protease to inhibitor, then, is very important in tissue homeostasis (reviewed in Yong, V.W. et al. (1998) Trends Neurosci. 21:75).

MMPs are implicated in a number of diseases including osteoarthritis (Mitchell, P. et al. (1996) J. Clin. Invest. 97:761), atherosclerotic plaque rupture (Sukhova, G.K. et al. (1999) Circulation 99:2503), aortic aneurysm (Schneiderman, J. et al. (1998) Am. J. Path. 152:703), non-healing wounds (Saarialho-Kere, U.K. et al. (1994) J. Clin. Invest. 94:79), bone resorption (Blavier, L. and J.M. Delaisse (1995) J. Cell Sci. 108:3649), age-related macular degeneration (Steen, B. et al. (1998) Invest. Ophthalmol. Vis. Sci. 39:2194), emphysema (Finlay, G.A. et al. (1997) Thorax 52:502), myocardial infarction (Rohde, L.E. et al. (1999) Circulation 99:3063) and dilated cardiomyopathy (Thomas, C.V. et al. (1998) Circulation 97:1708). MMP inhibitors prevent metastasis of mammary carcinoma and experimental tumors in rat, and Lewis lung carcinoma, hemangioma, and human ovarian carcinoma xenografts in mice (Eccles, S.A. et al. (1996) Cancer Res. 56:2815; Anderson et al. (1996) Cancer Res. 56:715-718; Volpert, O.V. et al. (1996) J. Clin. Invest. 98:671; Taraboletti, G. et al. (1995) J. NCI 87:293; Davies, B. et al. (1993) Cancer Res. 53:2087). MMPs may be active in Alzheimer's disease. A number of MMPs are implicated in multiple sclerosis, and administration of MMP inhibitors can relieve some of its symptoms (reviewed in Yong, supra).

Another family of metalloproteases is the ADAMs, for A Disintegrin and Metalloprotease Domain, which they share with their close relatives the adamalysins, snake venom metalloproteases (SVMPs). ADAMs combine features of both cell surface adhesion molecules and proteases, containing a prodomain, a protease domain, a disintegrin domain, a cysteine rich domain, an epidermal growth factor repeat, a transmembrane domain, and a cytoplasmic tail. The first three domains listed above are also found in the SVMPs. The ADAMs possess four potential functions: proteolysis, adhesion, signaling and fusion. The ADAMs share the metzincin zinc binding sequence and are inhibited by some MMP antagonists such as TIMP-1.

ADAMs are implicated in such processes as sperm-egg binding and fusion, myoblast fusion, and protein-ectodomain processing or shedding of cytokines, cytokine receptors, adhesion proteins and other extracellular protein domains (Schlöndorff, J. and C.P. Blobel (1999) J. Cell. Sci. 112:3603-3617). The Kuzbanian protein cleaves a substrate in the NOTCH pathway (possibly NOTCH itself), activating the program for lateral inhibition in Drosophila neural development. Two ADAMs, TACE (ADAM 17) and ADAM 10, are proposed to have analogous roles in the processing of amyloid precursor protein in the brain (Schlöndorff and Blobel, supra). TACE has also been identified as the TNF activating enzyme (Black, R.A. et al. (1997) Nature 385:729). TNF is a pleiotropic cytokine that is important in mobilizing host defenses in response to infection or trauma, but can cause severe damage in excess and is often overproduced in autoimmune disease. TACE cleaves membrane-bound pro-TNF to release a soluble form. Other ADAMs may be involved in a similar type of processing of other membrane-bound molecules.

The ADAMTS sub-family has all of the features of ADAM family metalloproteases and contain an additional thrombospondin domain (TS). The prototypic ADAMTS was identified in mouse, found to be expressed in heart and kidney and upregulated by proinflammatory stimuli (Kuno, K. et al. (1997) J. Biol. Chem. 272:556). To date eleven members are recognized by the Human Genome Organization (HUGO; http://www.gene.ucl.ac.uk/users/hester/adamts.html#Approved). Members of this family have the ability to degrade aggrecan, a high molecular weight proteoglycan which provides cartilage with important mechanical properties including compressibility, and which is lost during the development of arthritis. Enzymes which degrade aggrecan are thus considered attractive targets to prevent and slow the degradation of articular cartilage (See, e.g., Tortorella, M.D. (1999) Science 284:1664; Abbaszade, I. (1999) J. Biol. Chem. 274:23443). Other members are reported to have antiangiogenic potential (Kuno et al., supra) and/or procollagen processing (Colige, A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2374).

Protease inhibitors

Protease inhibitors and other regulators of protease activity control the activity and effects of proteases. Protease inhibitors have been shown to control pathogenesis in animal models of proteolytic disorders (Murphy, G. (1991) Agents Actions Suppl. 35:69-76). Low levels of the cystatins, low molecular weight inhibitors of the cysteine proteases, correlate with malignant progression of tumors (Calkins, C. et al. (1995) Biol. Biochem. Hoppe Seyler 376:71-80). Serpins are inhibitors of mammalian plasma serine proteases. Many serpins serve to regulate the blood clotting cascade and/or the complement cascade in mammals. Sp32 is a positive regulator of the mammalian acrosomal protease, acrosin, that binds the proenzyme, proacrosin, and thereby aides in packaging the enzyme into the acrosomal matrix (Baba, T. et al. (1994) J. Biol. Chem. 269:10133-10140). The Kunitz family of serine protease inhibitors are characterized by one or more "Kunitz domains"

containing a series of cysteine residues that are regularly spaced over approximately 50 amino acid residues and form three intrachain disulfide bonds. Members of this family include aprotinin, tissue factor pathway inhibitor (TFPI-1 and TFPI-2), inter-α-trypsin inhibitor, and bikunin. (Marlor, C.W. et al. (1997) J. Biol. Chem. 272:12202-12208.) Members of this family are potent inhibitors (in the nanomolar range) against serine proteases such as kallikrein and plasmin. Aprotinin has clinical utility in reduction of perioperative blood loss.

The discovery of new protein modification and maintenance molecules, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of protein modification and maintenance molecules.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, protein modification and maintenance molecules, referred to collectively as "PMMM" and individually as "PMMM-1," "PMMM-2," "PMMM-3," "PMMM-4," "PMMM-5," "PMMM-6," "PMMM-7," "PMMM-8," "PMMM-9," "PMMM-10," "PMMM-11," "PMMM-12," "PMMM-13," "PMMM-14," "PMMM-15," and "PMMM-16." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-16.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of

SEQ ID NO:1-16. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:17-32.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide

complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-

16. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PMMM, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PMMM, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PMMM, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide

having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions

whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5. Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing

particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"PMMM" refers to the amino acid sequences of substantially purified PMMM obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PMMM. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PMMM either by directly interacting with PMMM or by acting on components of the biological pathway in which PMMM participates.

An "allelic variant" is an alternative form of the gene encoding PMMM. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PMMM include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PMMM or a polypeptide with at least one functional characteristic of PMMM. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PMMM, and improper or unexpected hybridization to allelic

variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PMMM. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PMMM. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PMMM is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PMMM. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PMMM either by directly interacting with PMMM or by acting on components of the biological pathway in which PMMM participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PMMM polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that

makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PMMM, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PMMM or fragments of PMMM may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala

His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of PMMM or the polynucleotide encoding PMMM which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment

used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:17-32 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:17-32, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:17-32 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:17-32 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:17-32 and the region of SEQ ID NO:17-32 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-16 is encoded by a fragment of SEQ ID NO:17-32. A fragment of SEQ ID NO:1-16 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-16. For example, a fragment of SEQ ID NO:1-16 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-16. The precise length of a fragment of SEQ ID NO:1-16 and the region of SEQ ID NO:1-16 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of

molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62
Reward for match: 1
Penalty for mismatch: -2
Open Gap: 5 and Extension Gap: 2 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 11
Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode

similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for

chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such

similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PMMM which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PMMM which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PMMM. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PMMM.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which

comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an PMMM may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PMMM.

"Probe" refers to nucleic acid sequences encoding PMMM, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the

PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent,

chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PMMM, nucleic acids encoding PMMM, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term

"transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having

at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human protein modification and maintenance molecules (PMMM), the polynucleotides encoding PMMM, and the use of these compositions for the diagnosis, treatment, or prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites and potential glycosylation sites as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI), and amino acid residues comprising signature sequences, domains, and motifs. Column 5 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical

methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are protein modification and maintenance molecules.

For example, SEQ ID NO:1 is 56% identical from residue M1 to residue A16, 60% identical from residue C24 to residue Q76, and 53% identical, from residue G60 to residue A268, to Mus musculus tryptase 4 (GenBank ID g10947096) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.1e-78, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a trypsin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a serine protease.

As another example, SEQ ID NO:2 is 73% identical, from residue M1 to residue V379, to monkey prochymosin (GenBank ID g7008025) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.3e-142, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains an eukaryotic aspartyl protease domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:2 is an aspartic protease.

As another example, SEQ ID NO:6 is 60% identical, from residue S31 to residue H1120, to human zinc metalloendopeptidase ADAMTS10 (GenBank ID g11493589) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a reprolysin family propeptide, a reprolysin (M12B) family zinc metallopeptidase domain, and thrombospondin type 1 domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:6 is a zinc metalloprotease.

As another example, SEQ ID NO:7 is 41% identical, from residue L10 to residue N298, to an epidermis specific serine protease from Xenopus laevis (GenBank ID g6009515) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 8.7e-57, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 also contains a trypsin domain as determined by searching for statistically

significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:7 is a serine protease.

As another example, SEQ ID NO:8 is 44% identical, from residue R20 to residue M425, to human serine protease (GenBank ID g6137097) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.2e-87, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 also contains a SEA domain and a Trypsin site as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:8 is a serine protease (note that the "SEA domain" is found in enterokinase, a protease which cleaves the acidic propeptide from trypsinogen to yield active trypsin, (Kitamoto, Y. et al., (1994) Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592) and serine proteases from the trypsin family provide catalytic activity).

As another example, SEQ ID NO:11 is 32% identical, from residue C588 to residue S903, to Mus musculus bone morphogenetic protein (GenBank ID g439607) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.1e-62, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:11 also contains a CUB domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS, and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:11 is a developmentally regulated protease.

As another example, SEQ ID NO:12 is 43% identical (over 204 amino acid residues) to a murine thrombospondin type 1 domain (GenBank ID g4519541), characteristic of the ADAMTS metalloproteinases family, as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.4e-49, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also shares 30% identity (over 183 amino acid residues) with a Spodoptera frugiperda endoprotease (GenBank ID g1167860), with a BLAST probability score of 7.3e-10.

As another example, SEQ ID NO:13 is 37% identical (over 457 amino acid residues) to a human zinc metallopeptidase (GenBank ID g11493589), as determined by BLAST analysis, with a probability score is 4.5e-75. SEQ ID NO:13 also shares 34% identity (over 475 amino acid residues) with murine papilin (GenBank ID g11935122), a protease with homology to the ADAMTS metalloprotease family. The BLAST probability score is 5.9e-74. SEQ ID NO:13 also contains a thrombospondin type 1 domain as determined by searching for statistically significant matches in the

hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.)

As another example, SEQ ID NO:16 is 100% identical, from residue P119 to residue S365, to human bK57G9.1 (novel Kringle and CUB domain protein) (GenBank ID g6572252) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.2e-135, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains a CUB, a WSC, and a Kringle domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:16 is a protease. SEQ ID NO:3-5, SEQ ID NO:9-10, and SEQ ID NO:14-15 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-16 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:17-32 or that distinguish between SEQ ID NO:17-32 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as

FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors

which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PMMM variants. A preferred PMMM variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PMMM amino acid sequence, and which contains at least one functional or structural characteristic of PMMM.

The invention also encompasses polynucleotides which encode PMMM. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:17-32, which encodes PMMM. The polynucleotide sequences of SEQ ID NO:17-32, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PMMM. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PMMM. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:17-32 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:17-32. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PMMM.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding PMMM. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding PMMM, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding PMMM over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding PMMM. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PMMM.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PMMM, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be

produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PMMM, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PMMM and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PMMM under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PMMM or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PMMM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PMMM and PMMM derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PMMM or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:17-32 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA

sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PMMM may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the

emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PMMM may be cloned in recombinant DNA molecules that direct expression of PMMM, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PMMM.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PMMM-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of PMMM, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PMMM may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic

Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PMMM itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PMMM, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PMMM, the nucleotide sequences encoding PMMM or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PMMM. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PMMM. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PMMM and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PMMM and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and

16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PMMM. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PMMM. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PMMM can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PMMM into the vector's multiple cloning site disrupts the *lac*Z gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PMMM are needed, e.g. for the production of antibodies, vectors which direct high level expression of PMMM may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PMMM. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable

integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PMMM. Transcription of sequences encoding PMMM may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PMMM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PMMM in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PMMM in cell lines is preferred. For example, sequences encoding PMMM can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type..

Any number of selection systems may be used to recover transformed cell lines. These

include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PMMM is inserted within a marker gene sequence, transformed cells containing sequences encoding PMMM can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PMMM under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PMMM and that express PMMM may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PMMM using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PMMM is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana

Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PMMM include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PMMM, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PMMM may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PMMM may be designed to contain signal sequences which direct secretion of PMMM through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PMMM may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PMMM protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PMMM activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST),

maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PMMM encoding sequence and the heterologous protein sequence, so that PMMM may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PMMM may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

PMMM of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PMMM. At least one and up to a plurality of test compounds may be screened for specific binding to PMMM. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PMMM, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PMMM binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PMMM, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PMMM or cell membrane fractions which contain PMMM are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PMMM or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PMMM, either in solution or affixed to a solid support, and detecting the binding of PMMM to the compound.

Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PMMM of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PMMM. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PMMM activity, wherein PMMM is combined with at least one test compound, and the activity of PMMM in the presence of a test compound is compared with the activity of PMMM in the absence of the test compound. A change in the activity of PMMM in the presence of the test compound is indicative of a compound that modulates the activity of PMMM. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PMMM under conditions suitable for PMMM activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PMMM may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PMMM or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PMMM may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al.

(1998) Science 282:1145-1147).

Polynucleotides encoding PMMM can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PMMM is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PMMM, e.g., by secreting PMMM in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PMMM and protein modification and maintenance molecules. In addition, the expression of PMMM is closely associated with bone tumor, kidney, ovarian tumor, gastrointestinal, diseased prostate, uterus tumor, and brain tissue, including posterior cingulate tissue, as well as fibroblasts. Therefore, PMMM appears to play a role in gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders. In the treatment of disorders associated with increased PMMM expression or activity, it is desirable to decrease the expression or activity of PMMM. In the treatment of disorders associated with decreased PMMM expression or activity, it is desirable to increase the expression or activity of PMMM.

Therefore, in one embodiment, PMMM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PMMM. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-

antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism,

Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down

syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia.

In another embodiment, a vector capable of expressing PMMM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PMMM including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PMMM in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PMMM including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PMMM may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PMMM including, but not limited to, those listed above.

In a further embodiment, an antagonist of PMMM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PMMM. Examples of such disorders include, but are not limited to, those gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders described above. In one aspect, an antibody which specifically binds PMMM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PMMM.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PMMM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PMMM including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PMMM may be produced using methods which are generally known in the art. In particular, purified PMMM may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PMMM. Antibodies to PMMM may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with PMMM or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PMMM have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PMMM amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PMMM may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J.

Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PMMM-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PMMM may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PMMM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PMMM epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PMMM. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of PMMM-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PMMM epitopes, represents the average affinity, or avidity, of the antibodies for PMMM. The K_a determined for a preparation of monoclonal antibodies, which are monospecific

for a particular PMMM epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the PMMM-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PMMM, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PMMM-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PMMM, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PMMM. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PMMM. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PMMM may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PMMM expression or regulation causes disease, the expression of PMMM from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PMMM are treated by constructing mammalian expression vectors encoding PMMM and introducing these vectors by mechanical means into PMMM-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of PMMM include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PMMM may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and

H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PMMM from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PMMM expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PMMM under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PMMM to cells which have one or more genetic abnormalities with respect

to the expression of PMMM. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PMMM to target cells which have one or more genetic abnormalities with respect to the expression of PMMM. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PMMM to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PMMM to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PMMM into the

alphavirus genome in place of the capsid-coding region results in the production of a large number of PMMM-coding RNAs and the synthesis of high levels of PMMM in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PMMM into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PMMM.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis.

Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PMMM. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PMMM. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PMMM expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PMMM may be therapeutically useful, and in the treatment of disorders associated with decreased PMMM expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PMMM may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PMMM is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PMMM are assayed

by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PMMM. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PMMM, antibodies to PMMM, and mimetics, agonists, antagonists, or inhibitors of PMMM.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PMMM or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PMMM or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PMMM or fragments thereof, antibodies of PMMM, and agonists, antagonists or inhibitors of PMMM, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind PMMM may be used for the diagnosis of disorders characterized by expression of PMMM, or in assays to monitor patients being treated with PMMM or agonists, antagonists, or inhibitors of PMMM. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PMMM include methods which utilize the antibody and a label to detect PMMM in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PMMM, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PMMM expression. Normal or standard values for PMMM expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PMMM under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PMMM expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PMMM may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect

and quantify gene expression in biopsied tissues in which expression of PMMM may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PMMM, and to monitor regulation of PMMM levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PMMM or closely related molecules may be used to identify nucleic acid sequences which encode PMMM. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PMMM, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PMMM encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:17-32 or from genomic sequences including promoters, enhancers, and introns of the PMMM gene.

Means for producing specific hybridization probes for DNAs encoding PMMM include the cloning of polynucleotide sequences encoding PMMM or PMMM derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PMMM may be used for the diagnosis of disorders associated with expression of PMMM. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-

occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone

resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental

retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding PMMM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PMMM expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PMMM may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PMMM may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PMMM in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PMMM, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PMMM, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from

normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PMMM may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PMMM, or a fragment of a polynucleotide complementary to the polynucleotide encoding PMMM, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PMMM may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PMMM are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the

sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of PMMM include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and

effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PMMM, fragments of PMMM, or antibodies specific for PMMM may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of

gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PMMM to quantify

the levels of PMMM expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are

well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PMMM may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PMMM on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PMMM, its catalytic or immunogenic fragments, or

oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PMMM and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PMMM, or fragments thereof, and washed. Bound PMMM is then detected by methods well known in the art. Purified PMMM can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PMMM specifically compete with a test compound for binding PMMM. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PMMM.

In additional embodiments, the nucleotide sequences which encode PMMM may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/269,581, U.S. Ser. No. 60/271,198, U.S. Ser. No. 60/272,813, U.S. Ser. No. 60/278,505, U.S. Ser. No. 60/280,539, U.S. Ser. No. 60/266,762, U.S. Ser. No. 60/265,705, and U.S. Ser. No. 60/275,586, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine

isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal

cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open

reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:17-32. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative protein modification and maintenance molecules were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode protein modification and maintenance molecules, the encoded polypeptides were analyzed by querying against PFAM models for protein modification

and maintenance molecules. Potential protein modification and maintenance molecules were also identified by homology to Incyte cDNA sequences that had been annotated as protein modification and maintenance molecules. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of PMMM Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:17-32 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:17-32 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:30 was mapped to chromosome 5 within the interval from 174.30 centiMorgans to the q terminus, and to chromosome 10 within the interval from 83.30 to

96.90 centiMorgans. More than one map location is reported for SEQ ID NO:30, indicating that sequences having different map locations were assembled into a single cluster. This situation occurs, for example, when sequences having strong similarity, but not complete identity, are assembled into a single cluster.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PMMM are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is

classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PMMM. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of PMMM Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar,

Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in PMMM Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:17-32 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice

variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:17-32 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc.

(CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash

buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each

spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XII. Complementary Polynucleotides

Sequences complementary to the PMMM-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PMMM. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PMMM. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PMMM-encoding transcript.

XIII. Expression of PMMM

Expression and purification of PMMM is achieved using bacterial or virus-based expression systems. For expression of PMMM in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PMMM upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PMMM in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PMMM by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PMMM is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on

immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PMMM at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PMMM obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functional Assays

PMMM function is assessed by expressing the sequences encoding PMMM at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PMMM on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PMMM and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake

Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PMMM and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of PMMM Specific Antibodies

PMMM substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the PMMM amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PMMM activity by, for example, binding the peptide or PMMM to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring PMMM Using Specific Antibodies

Naturally occurring or recombinant PMMM is substantially purified by immunoaffinity chromatography using antibodies specific for PMMM. An immunoaffinity column is constructed by covalently coupling anti-PMMM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PMMM are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PMMM (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PMMM binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PMMM is collected.

XVII. Identification of Molecules Which Interact with PMMM

PMMM, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PMMM, washed,

and any wells with labeled PMMM complex are assayed. Data obtained using different concentrations of PMMM are used to calculate values for the number, affinity, and association of PMMM with the candidate molecules.

Alternatively, molecules interacting with PMMM are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PMMM may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of PMMM Activity

Protease activity is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases, or metalloproteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (carboxypeptidases A and B, procollagen C-proteinase). Commonly used chromogens are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. Assays are performed at ambient temperature and contain an aliquot of the enzyme and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette, and the increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate is measured. The change in absorbance is proportional to the enzyme activity in the assay.

An alternate assay for ubiquitin hydrolase activity measures the hydrolysis of a ubiquitin precursor. The assay is performed at ambient temperature and contains an aliquot of PMMM and the appropriate substrate in a suitable buffer. Chemically synthesized human ubiquitin-valine may be used as substrate. Cleavage of the C-terminal valine residue from the substrate is monitored by capillary electrophoresis (Franklin, K. et al. (1997) Anal. Biochem. 247:305-309).

In the alternative, an assay for protease activity takes advantage of fluorescence resonance energy transfer (FRET) that occurs when one donor and one acceptor fluorophore with an appropriate spectral overlap are in close proximity. A flexible peptide linker containing a cleavage site specific for PMMM is fused between a red-shifted variant (RSGFP4) and a blue variant (BFP5) of Green Fluorescent Protein. This fusion protein has spectral properties that suggest energy transfer is occurring from BFP5 to RSGFP4. When the fusion protein is incubated with PMMM, the substrate is cleaved, and the two fluorescent proteins dissociate. This is accompanied by a marked decrease in

energy transfer which is quantified by comparing the emission spectra before and after the addition of PMMM (Mitra, R.D. et al. (1996) Gene 173:13-17). This assay can also be performed in living cells. In this case the fluorescent substrate protein is expressed constitutively in cells and PMMM is introduced on an inducible vector so that FRET can be monitored in the presence and absence of PMMM (Sagot, I. et al. (1999) FEBS Lett. 447:53-57).

XVIII. Identification of PMMM Substrates

Phage display libraries can be used to identify optimal substrate sequences for PMMM. A random hexamer followed by a linker and a known antibody epitope is cloned as an N-terminal extension of gene III in a filamentous phage library. Gene III codes for a coat protein, and the epitope will be displayed on the surface of each phage particle. The library is incubated with PMMM under proteolytic conditions so that the epitope will be removed if the hexamer codes for a PMMM cleavage site. An antibody that recognizes the epitope is added along with immobilized protein A. Uncleaved phage, which still bear the epitope, are removed by centrifugation. Phage in the supernatant are then amplified and undergo several more rounds of screening. Individual phage clones are then isolated and sequenced. Reaction kinetics for these peptide substrates can be studied using an assay in Example XVII, and an optimal cleavage sequence can be derived (Ke, S.H. et al. (1997) J. Biol. Chem. 272:16603-16609).

To screen for in vivo PMMM substrates, this method can be expanded to screen a cDNA expression library displayed on the surface of phage particles (T7SELECT 10-3 Phage display vector, Novagen, Madison WI) or yeast cells (pYD1 yeast display vector kit, Invitrogen, Carlsbad CA). In this case, entire cDNAs are fused between Gene III and the appropriate epitope.

XIX. Identification of PMMM Inhibitors

Compounds to be tested are arrayed in the wells of a multi-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. PMMM activity is measured for each well and the ability of each compound to inhibit PMMM activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PMMM activity.

In the alternative, phage display libraries can be used to screen for peptide PMMM inhibitors. Candidates are found among peptides which bind tightly to a protease. In this case, multi-well plate wells are coated with PMMM and incubated with a random peptide phage display library or a cyclic peptide library (Koivunen, E. et al. (1999) Nat. Biotechnol. 17:768-774). Unbound phage are washed away and selected phage amplified and rescreened for several more rounds. Candidates are tested for PMMM inhibitory activity using an assay described in Example XVIII.

Various modifications and variations of the described methods and systems of the invention

will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7482256	1	7482256CD1	17	7482256CB1
71973513	2	71973513CD1	18	71973513CB1
7648238	3	7648238CD1	19	7648238CB1
1719204	4	1719204CD1	20	1719204CB1
7472647	5	7472647CD1	21	7472647CB1
7472654	6	7472654CD1	22	7472654CB1
7480224	7	7480224CD1	23	7480224CB1
7481056	8	7481056CD1	24	7481056CB1
3750264	9	3750264CD1	25	3750264CB1
1749735	10	1749735CD1	26	1749735CB1
7473634	11	7473634CD1	27	7473634CB1
4767844	12	4767844CD1	28	4767844CB1
7487584	13	7487584CD1	29	7487584CB1
1468733	14	1468733CD1	30	1468733CB1
1652084	15	1652084CD1	31	1652084CB1
3456896	16	3456896CD1	32	3456896CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7482256CD1	g10947096	3.1E-78	[Mus musculus] tryptase 4
2	71973513CD1	g7008025	4.3E-142	[Callithrix jacchus] prochymosin Kageyama,T. (2000) J. Biochem. (Tokyo) 127:761-770
3	7648238CD1	g4323041	9.1E-46	[Homo sapiens] caspase 14 precursor
4	1719204CD1	g1865716	0.0	[Bos taurus] procollagen I N-proteinase
5	7472647CD1	g15099921	0.0	[Homo sapiens] ADAM-TS related protein 1
		g11935122	7.9E-88	[Mus musculus] papilin Kramerova, I.A., (2000) Development 127:5475-5485 Papilin in development; a pericellular protein with a homology to the ADAMTS metalloproteinases.
6	7472654CD1	g11493589	0.0	[5' incom][Homo sapiens] zinc metalloendopeptidase
7	7480224CD1	g6009515	8.7E-57	[Xenopus laevis] epidermis specific serine protease
8	7481056CD1	g6137097	2.2E-87	[Homo sapiens] serine protease DESC1
9	3750264CD1	g11493589	0.0	[Homo sapiens] zinc metalloendopeptidase Hurskainen,T.L., et al., (1999) J. Biol. Chem. 274:25555-25563
11	7473634CD1	g10185056	1.4E-62	[Gallus gallus] colloid protein Liaubet, L. et al. (2000) Mech. Dev. 96:101-105
		g439607	1.1E-62	[Mus musculus] bone morphogenetic protein Fukagawa, M. et al. (1994) Dev. Biol. 163:175-183
12	4767844CD1	g4519541	9.4E-49	[Mus musculus] thrombospondin type 1 domain
13	7487584CD1	g15099921	0.0	[Homo sapiens] ADAM-TS related protein 1
		g11493589	4.5E-75	[Homo sapiens] zinc metalloendopeptidase

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
14	1468733CD1	g35328	5.7E-140	[Homo sapiens] protease small subunit (aa 1-268) Ohno, S. et.al. (1986) Nucleic Acids Res. 14:5559 Nucleotide sequence of a cDNA coding for the small subunit of human calcium-dependent protease. ; Zhang,W. et al. (1996) J. Biol. Chem. 271:18825-18830 The major calpain isozymes are long-lived proteins. Design of an antisense strategy for calpain depletion in cultured cells.
15	1652084CD1	g16226029	0.0	[Homo sapiens] serine proteinase inhibitor SERPINB11
		g164241	4E-84	[Equus caballus] serpin Kordula,T. et al. (1993) Biochem. J. 293 (Pt 1): 187-193 Molecular cloning and expression of an intracellular serpin: an elastase inhibitor from horse leucocytes.
		g16226021	0.0	[Homo sapiens] serine proteinase inhibitor SERPINB11
16	3456896CD1	g6572252	1.2E-135	bK57G9.1 (novel Kringle and CUB domain protein) [Homo sapiens]

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites, Potential Glycosylation Sites, Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7482256	269	Signal_Peptide: M1-G19	SPSCAN
			Signal Peptide: M1-G25	HMMER
			Trypsin: V33-I243	HMMER_PFAM
			Kringle domain proteins. BL00021: C58-F75, I117-G138, G202-I243	BLIMPS_BLOCKS
			Serine proteases, trypsin BL00134: C58-C74, D194-I217, P230-I243	BLIMPS_BLOCKS
			Apple (serine protease) domain proteins BL00495: L69-S107, V108-P142, A186-W220	BLIMPS_BLOCKS
			Serine proteases, trypsin family, active sites ; trypsin_his.prf: L50-A100; trypsin_ser.prf: I179-Q226	PROFILESCAN
			Chymotrypsin serine protease family (S1) signature PR00722: G59-C74, V94-V108, V193-V205	BLIMPS_PRINTS
			PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: V82-I243, V33-S78	BLAST_PRODOM
			TRYPSIN DM00018; P15944|31-270: F75-R245, V33-C74; Q02844|29-268: V82-I243, V33-C74 P15157|31-270: L62-I243, V33-C74; P21845|31-271: D98-R245, V33-C74	BLAST_DOMO
			Potential Phosphorylation Sites: S39 S49 S64 S174 T195 T251	MOTIFS
			Potential Glycosylation Sites: N162 N235	MOTIFS
			Serine proteases, trypsin family, histidine active site L69-C74	MOTIFS
			Serine proteases, trypsin family, serine active site D194-V205	MOTIFS
2	71973513	379	Signal_cleavage: M1-A18	SPSCAN
			Signal Peptide: M1-N17, M1-T20	HMMER
			Eukaryotic aspartyl protease: S65-E190, R198-A378	HMMER_PFAM
			Tranmembrane domains: M1-S29, L243-C263; N terminus is cytosolic.	TMAP
			Eukaryotic and viral aspartyl proteases proteins BL00141: F87-S102, D177-A188, R208-G217, A269-L278, I353-A376	BLIMPS_BLOCKS
			Pepsin (A1) aspartic protease family signature; PR00792: I80-V100, S203-T216, A269-G280, W352-D367	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites, Potential Glycosylation Sites, Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEASE ASPARTYL HYDROLASE PRECURSOR SIGNAL ZYMOGEN GLYCOPROTEIN ASPARTIC PROTEINASE MULTIGENE; PD000182: S119-A378, L66-S189	BLAST_PRODOM
			EUKARYOTIC AND VIRAL ASPARTYL PROTEASES ; DM00126|P00794|18-379: I19-A378; DM00126|P16476|16-381: I19-A378 DM00126|P03954|16-386: I19-A376; DM00126|P28713|16-385: I19-A378	BLAST_DOMO
			Potential Phosphorylation Sites: S29 S52 S56 S138 S163 S174 S364 T172 T206 T225 T332 Y214	MOTIFS
			Eukaryotic and viral aspartyl proteases active site: L89-V100, A269-G280	MOTIFS
3	7648238	398	ICE-like protease (caspase) p10 domain: A308-V366; p20 domain: R269-A292, R183-F222	HMMER_PFAM
			Caspase family histidine proteins BL01121: I180-F215, C229-G244, C270-G287, S311-E345, L359-V371	BLIMPS_BLOCKS
			Interleukin-1B converting enzyme signature PR00376: R183-G201, G201-L219, A236-G244, C270-G288	BLIMPS_PRINTS
			INTERLEUKIN-1 BETA CONVERTING ENZYME FAMILY HISTIDINE DM01067|P42576|136-311: I180-G288; DM01067|P29594|149-323: I180-V294	BLAST_DOMO
			Potential Phosphorylation Sites: S91 S141 S314 S389 T13 T164 T205 T228 T342	MOTIFS
4	1719204	1221	Signal Peptide: M1-A22, M1-S24, M1-E28	HMMER
			Signal Cleavage: M1-G23	SPSCAN
			Reprolysin family propeptide domain: R120-V240	HMMER_PFAM
			Reprolysin (M12B) family zinc metallopeptidase domain: I261-P460	HMMER_PFAM
			Thrombospondin type 1 domain: A968-C1019, S556-C604, Y847-C904, W909-C966	HMMER_PFAM
			Transmembrane domains: P3-A21 L300-Y316; N-terminus is cytosolic	TMAP
			Neutral zinc metallopeptidases signature BL00142: V395-G405	BLIMPS_BLOCKS
			PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE C02B4.1 A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD013511: L471-E546; PD011654: Q642-C711	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites, Potential Glycosylation Sites, Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN F25H8.3 F53B6.2 KIAA0605 PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR ALTERNATIVE; PD007018: W849-Q969, W909-C1019	BLAST_PRODOM
			PROCOLLAGEN I NPROTEINASE EC 3.4.24.14 PROCOLLAGEN NENDOPEPTIDASE HYDROLASE; PD132243: Q1041-P1171	BLAST_PRODOM
			ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN; DM00368|Q05910|189-395: I261-P460; DM00368|A42972|5-205: I261-P460 DM00368|JC2550|1-201: I261-P460; DM00368|P20164|1-203: P256-P460	BLAST_DOMO
			Potential Phosphorylation Sites: S32 S132 S169 S200 S321 S348 S442 S477 S508 S621 S670 S694 S793 S1056 S1096 T247 T360 T518 T607 T713 T772 T941 T981 T1027 T1136 Y549	MOTIFS
			Potential Glycosylation Sites: N109 N475 N939 N1025	MOTIFS
5	7472647	1537	Signal Peptide: M1-S28	HMMER
			Signal Cleavage: M1-S28	SPSCAN
			Immunoglobulin domain: G1076-A1130, K667-A724, ; G1186-A1246, S972-A1027	HMMER_PFAM
			Thrombospondin type 1 domain: D37-C81, F526-C583, S1322-C1382, W440-C492, W380-C437, V1443-C1500	HMMER_PFAM
			Transmembrane domains: C4-R27 R650-R678 V1213-A1232; N-terminus is cytosolic	TMAP
			PROTEIN F25H8.3 F53B6.2 KIAA0605 PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR ALTERNATIVE; PD007018: W1265-C1382	BLAST_PRODOM
			PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1; PD011654: P115-C185	BLAST_PRODOM
			Potential Phosphorylation Sites: S22 S28 S56 S62 S77 S120 S252 S329 S402 S414 S475 S558 S574 S631 S748 S751 S781 S794 S829 S886 S898 S903 S919 S924 S932 S946 S952 S999 S1119 S1127 S1238 S1464 T8 T25 T169 T184 T199 T235 T320 T413 T423 T648 T769 T827 T828 T940 T1050 T1058 T1070 T1153 T1342 T1346 T1474 T1498 T1508 Y226 Y720	MOTIFS
			Potential Glycosylation Sites: N251 N779 N826 N859 N1026 N1078 N1098 N1117 N1202 N1233 N1293	MOTIFS
6	7472654	1120	Signal Peptide: M1-S23	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites, Potential Glycosylation Sites, Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Signal Cleavage: M1-S23	SPSCAN
			Reprolysin family propeptide: N99-H206	HMMER_PFAM
			Reprolysin (M12B) family zinc metallopeptidase domain: R250-P468	HMMER_PFAM
			Thrombospondin type 1 domain: G562-C615, G909-C962, W847-C902, W966-C1020, W1025-C1075	HMMER_PFAM
			Neutral zinc metallopeptidases signature BL00142: T400-G410	BLIMPS_BLOCKS
			PROTEIN F25H8.3 F53B6.2 KIAA0605 PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR ALTERNATIVE; PD007018: W847-Q965, W966-C1075	BLAST_PRODOM
			METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION; PD000791: E249-P468	BLAST_PRODOM
			PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1; PD011654: C653-C719	BLAST_PRODOM
			ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN; DM00368|S48160|193-396: V294-P468; DM00368|S60257|204-414: H350-P468; DM00368|P22796|1-199: V295-P468; DM00368|P20164|1-203: V295-P468	BLAST_DOMO
			Neutral zinc metallopeptidases, zinc-binding region signature: T400-F409	MOTIFS
			Potential Phosphorylation Sites: S30 S31 S67 S72 S215 S388 S454 S458 S516 S581 S717 S764 S936 S1073 S1081 T37 T60 T143 T160 T173 T341 T357 T363 T462 T497 T666 T796 T948 T975 T1062 Y770	MOTIFS
			Potential Glycosylation Sites: N99 N172 N222 N234 N727 N959	MOTIFS
7	7480224	328	Signal peptide: M1-G20	SPScan
			Signal peptides: M1-Q21, M1-P22, M1-R27	HMMER
			Trypsin domain: V28-I262	HMMER-PFAM
			Serine proteases, trypsin family, active sites: L45-K93, I199-K246	ProfileScan
			Trypsin family serine proteases: histidine active site: L64-C69 serine active site D214-S225	MOTIFS
			Transmembrane domains: A4-R27, N271-S292; N-terminus is non-cytosolic	TMAP
			Serine proteases, trypsin BL00134: Y53-C69, D214-V237, P249-I262	BLIMPS-BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites, Potential Glycosylation Sites, Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Apple domain proteins BL00495: M1-W41, V124-E158, A206-W240, W240-R268	BLIMPS-BLOCKS
			Type I fibronectin BL01253: Y53-A66, S122-E158, D161-I199, K213-C226, V231-T265	BLIMPS-BLOCKS
			Chymotrypsin serine protease family (S1) signature PR00722: G54-C69, D110-V124, K213-S225	BLIMPS-PRINTS
			Serine protease PD000046: G54-I262	BLAST-PRODOM
			Trypsin DM00018: A57014|45-284: V28-I266 P21845|31-271: V28-N263 P15944|31-270: V28-N263 P15157|31-270: V28-N263	BLAST-DOMO
			Potential Phosphorylation Sites: S25 S59 S91 S160 S215 S324 T87 T111 T305 Y164 Y185	MOTIFS
			Potential Glycosylation Sites: N263	MOTIFS
8	7481056	425	SEA domain: D55-N181	HMMER_PFAM
			Trypsin: V194-I419	HMMER_PFAM
			Transmembrane domain: F24-V52; N-terminus is non-cytosolic	TMAP
			Kringle domain proteins. BL00021: C220-F237, V299-G320, G378-I419	BLIMPS_BLOCKS
			Serine proteases, trypsin BL00134: C220-C236, D370-I393, P406-I419	BLIMPS_BLOCKS
			Apple domain proteins. BL00495: S81-D119, S167-W207, A222-I254, G251-G289, V290-D324, A362-W396, G397-M425	BLIMPS_BLOCKS
			Serine proteases, trypsin family, active sites: Q212-N262	PROFILESCAN
			Serine proteases, trypsin family, active sites: I355-L402	PROFILESCAN
			Chymotrypsin serine protease family (S1) signature PR00722: G221-C236, T276-V290, I369-V381	BLIMPS_PRINTS
			PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: T288-I419	BLAST_PRODOM
			AIRWAY TRYPSINLIKE PROTEASE PROTEASE PD103718: Q23-T171	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites, Potential Glycosylation Sites, Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			TRYPSIN DM00018|P23578|42-289: R192-K422 DM00018|P05981|163-403: I193-I419 DM00018|P14272|391-624: I193-K422 DM00018|P10323|42-288: R192-K422	BLAST_DOMO
			Potential Phosphorylation Sites: S9 S14 S27 S64 S80 S117 S153 S167 S305 S321 T190 T199 T288 T331 Y151	MOTIFS
			Serine proteases, trypsin family, histidine active site: L231-C236	MOTIFS
			Serine proteases, trypsin family, serine active site: D370-V381	MOTIFS
9	3750264	1103	Signal_cleavage: M1-A25	SPSCAN
			Signal Peptide: M1-R27, M1-A25	HMMER
			Reprolysin family propeptide: N90-P201	HMMER_PFAM
			Reprolysin (M12B) family zinc metallo: R239-P457	HMMER_PFAM
			Thrombospondin type 1 domain: G551-C601, W829-C884, W1007-C1057, W888-C944, P946-C1002	HMMER_PFAM
			Transmembrane domain: A4-H24, S787-L808; N-terminus is non-cytosolic	TMAP
			PRECURSOR GLYCOPROTEIN S PD01719: W550-P577, R877-C884	BLIMPS_PRODOM
			PROTEIN F25H8.3 F53B6.2 KIAA0605 PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR ALTERNATIVE PD007018: W829-E947	BLAST_PRODOM
			PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: C639-C705	BLAST_PRODOM
			ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN; DM00368|S60257|204-414: N338-P457 DM00368|P28891|1-202: H339-P457 DM00368|P14530|1-201: N338-P457	BLAST_DOMO
			THROMBOSPONDIN TYPE 1 REPEAT DM00275|P35440|485-548: P543-C596	BLAST_DOMO
			Leucine zipper pattern L280-L301	MOTIFS
			Neutral zinc metallopeptidases, zinc-binding region signature T389-F398	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites, Potential Glycosylation Sites, Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S28 S34 S94 S170 S184 S377 S443 S505 S541 S570 S576 S614 S703 S916 S1027 T45 T68 T211 T224 T346 T425 T630 T652 T994 T1061	MOTIFS
			Potential Glycosylation Sites: N90 N222 N323 N740 N795 N892	MOTIFS
10	1749735	83	Signal_cleavage: M1-S16	SPSCAN
			Signal Peptide: M1-V21, M1-C20, M1-D25	HMMER
			Eukaryotic thiol (cysteine) proteases active site PDOC00126: S10-N83	PROFILESCAN
			Serine proteases, trypsin family, histidine active site L62-C67	MOTIFS
11	7473634	1274	Signal_cleavage: M1-S16	SPSCAN
			CUB domain: C623-Y728, C449-Y554, C276-Y384, C1142-F1248, C73-F174, C969-Y1074, C795-F902	HMMER_PFAM
			GLYCOPROTEIN DOMAIN EGF-LIKE PROTEIN PRECURSOR SIGNAL RECEPTOR INTRINSIC FACTOR B12 REPEAT PD000165: C73-V176, C623-Y728, C1142-F1248, T454-Y554, C271-Y384	BLAST_PRODOM
			COMPLEMENT REGULATORY PROTEIN PD060257: V1080-W1171	BLAST_PRODOM
			C1R/C1S REPEAT DM00162 I49540|748-862: E620-F724, C449-T555, E70-A172, A1140-S1249, C276-A382 I49540|592-708: C619-S730, C445-F550, C1138-F1248, E70-F174 P98063|755-862: L627-F724, T454-T555, T80-A172, A1149-S1249, S284-A382 A57190|826-947: V611-S730, C73-F174, P789-F902	BLAST_DOMO
			Potential Phosphorylation Sites: S54 S91 S130 S150 S196 S239 S353 S520 S660 S737 S771 S844 S856 S903 S919 S972 S987 S1031 S1064 S1151 S1260 T37 T76 T307 T309 T332 T546 T769 T872 T901 T1021 T1039 T1075 T1255 Y674	MOTIFS
			Potential Glycosylation Sites: N452 N551 N820 N880 N899 N1049 N1062	MOTIFS
			ATP/GTP-binding site motif A (P-loop): G796-S803	MOTIFS
			Glycosyl hydrolase family 10: G897-L907	MOTIFS
12	4767844	243	Signal cleavage: M1-C21	SPSCAN
			Signal Peptide: M1-G23	HMMER
			Potential Phosphorylation Sites: S29 S33 S193 T189 T199 T209 T238	MOTIFS
			Potential Glycosylation Sites: N160	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites, Potential Glycosylation Sites, Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	7487584	672	Signal cleavage: M1-S28	SPSCAN
			Signal Peptide: M1-E30	HMMER
			Thrombospondin type 1 domain: F526-C583, W440-C492, W380-C437, D37-C81, W611-C666	HMMER-PFAM
			TMAP: C4-R27; N-terminus is not cytoplasmic	TMAP
			PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1: PD011654: P115-C185	BLAST-PRODOM
			Potential Phosphorylation Sites: T8, S22, T25, S28, S56, S62, S77, S120, T169, T184, T199, Y226, T235, S252, T320, S329, S402, T413, S414, T423, S475, S558, S574, T650, S651	MOTIFS
			Potential Glycosylation Sites: N251	MOTIFS
14	1468733	442	EF hand: T317-I345, R347-A375, A412-T439, L383-L410	HMMER_PFAM
			RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain): V55-L123	HMMER_PFAM
			Transmembrane domains: A4-Q22, G191-G213, G227-E245; N terminus is non-cytosolic.	TMAP
			CALPAIN SUBUNIT CALCIUM-BINDING NEUTRAL PROTEASE CALCIUM ACTIVATED PROTEINASE CANP HYDROLASE LARGE; PD003609: E270-K339; PD002827: L341-I404	BLAST_PRODOM
			SMALL SUBUNIT CALPAIN CALCIUM DEPENDENT REGULATORY CALCIUM ACTIVATED NEUTRAL PROTEINASE CANP; PD015187: T231-S269	BLAST_PRODOM
			PROTEIN RNA-BINDING REPEAT NUCLEAR RIBO-NUCLEOPROTEIN HETEROGENEOUS ; PD150499: V55-L123	BLAST_PRODOM
			CALPAIN CATALYTIC DOMAIN ; DM01221|P13135|161-261: Y340-Y441; DM01221|P20807|719-819: Y340-Y441	BLAST_DOMO
			RIBONUCLEOPROTEIN REPEAT ; DM00012|P31943|284-363: Q48-T128; DM00012|P52597|284-363: Q48-T128	BLAST_DOMO
			Potential Phosphorylation Sites: S262 S290 S392 T39 T65 T101 T317 T330 T357 Y70 Y340	MOTIFS
			Potential Glycosylation Sites: N126 N146 N168 N267	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites, Potential Glycosylation Sites, Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			EF-hand calcium-binding domains: D326-F338, D356-L368	MOTIFS
15	1652084	378	Serpins (serine protease inhibitors): M1-P378	HMMER_PFAM
			Transmembrane domains: I24-A46, P223-L242; N terminus is cytosolic.	TMAP
			Serpins proteins; BL00284: N27-T50, T131-F151, S160-M201, V270-F296, N354-P378	BLIMPS_BLOCKS
			Serpins signature serpin: T330-P378	PROFILESCAN
			SERPIN INHIBITOR PROTEASE SERINE SIGNAL PRECURSOR GLYCOPROTEIN PLASMA PROTEIN PROTEINASE; PD000192: L4-P378	BLAST_PRODOM
			SERPINS ; DM00112|P05619|2-377: L4-S377; DM00112|P48595|2-395: K82-S377, S3-V57; DM00112|P01014|2-386: S3-K374; DM00112|S38962|23-376: N23-S377	BLAST_DOMO
			Potential Phosphorylation Sites: S72 S80 S109 S111 S127 S154 S321 T131 T183 T206 T253 Y281	MOTIFS
			Potential Glycosylation Sites: N59 N86 N141 N195	MOTIFS
			Serpins signature: F351-I361	MOTIFS
			Signal peptide: M1-G48	SPSCAN
16	3456896	458	Signal_cleavage: M1-A20	SPSCAN
			Signal PeptideS: M1-P22, M1-G27, M1-P24, M1-A20, M1-R21	HMMER
			CUB domain: C216-Y320	HMMER_PFAM
			WSC domain: N121-G202	HMMER_PFAM
			Kringle domain: C34-C116	HMMER_PFAM
			Transmembrane domains: P4-A20, H285-Q312, G375-K403; N terminus is cytosolic	TMAP
			Kringle domain signature and profile: N61-E112	PROFILESCAN
			Kringle domain signature PR00018: C34-T49, Q52-F64, G79-V99, G105-C116	BLIMPS_PRINTS
			PRECURSOR SIGNAL SERINE GLYCOPROTEIN PROTEASE KRINGLE HYDROLASE PLASMA GROWTH PLASMINOGEN; PD000395: C34-C116	BLAST_PRODOM
			KRINGLE ; DM00069|P00750|206-305: P22-G120; DM00069|P20918|263-357: P24-Q117; DM00069|P06868|244-338: P24-Q117; DM00069|P20918|359-460: E33-G120	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites, Potential Glycosylation Sites, Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S141 S155 S307 S355 S404 S447 T70 T137 T238 T245 T277 T337 T401 T421	MOTIFS
			Potential Glycosylation Sites: N47 N61 N219 N295 N335 N347	MOTIFS
			Kringle domain signature: Y85-D90	MOTIFS

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
17/7482256CB1/993	1-735, 592-706, 618-980, 822-927, 822-928, 822-993
18/71973513CB1/1238	1-1137, 1-1140, 62-213, 179-213, 448-564, 448-572, 448-573, 476-572, 510-572, 528-572, 592-705, 593-701, 860-1238, 886-1238, 902-1058, 902-1108, 902-1122, 902-1160, 902-1206, 902-1228, 902-1233, 902-1234, 902-1236, 902-1238, 936-1238
19/7648238CB1/1233	1-396, 74-600, 74-672, 107-792, 128-203, 136-802, 164-203, 167-889, 178-836, 203-547, 203-842, 204-725, 204-759, 204-935, 205-966, 206-885, 206-903, 207-547, 211-890, 216-909, 218-869, 236-710, 264-992, 264-1004, 268-846, 278-547, 283-779, 287-606, 289-869, 290-974, 299-987, 315-964, 322-849, 326-950, 397-1233, 411-926, 414-764, 435-1016, 450-809, 452-898, 469-962, 521-1015, 527-773, 527-1015, 527-1016, 543-1017, 589-935, 715-1015, 826-1003, 828-899
20/1719204CB1/5511	1-500, 83-245, 83-247, 118-623, 521-870, 592-1138, 608-1134, 608-1138, 653-1137, 653-1138, 871-3513, 1009-1754, 1302-2052, 1543-2052, 2172-2252, 2174-2252, 2242-2752, 2276-2935, 2683-3265, 2724-3241, 2750-3304, 2837-3333, 2985-3633, 3002-3586, 3130-3869, 3131-3869, 3161-3869, 3173-3429, 3173-3869, 3179-3869, 3195-3951, 3213-3951, 3321-3972, 3375-4163, 3378-3709, 3383-3869, 3450-3720, 3550-4201, 3631-4247, 3634-4224, 3807-4070, 3807-4078, 3807-4082, 3807-4097, 3807-4239, 3807-4288, 3807-4358, 3807-4394, 3838-4075, 3838-4270, 3861-4472, 3960-4317, 3971-4487, 4171-4449, 4173-4443, 4173-4654, 4174-4470, 4174-4760, 4208-4466, 4251-4655, 4305-4541, 4305-4670, 4305-4859, 4382-5211, 4406-4621, 4406-4684, 4421-4678, 4433-5211, 4472-5262, 4517-5260, 4523-5248, 4561-5222, 4566-5174, 4583-4815, 4583-5130, 4591-5258, 4593-4900, 4593-5174, 4597-5244, 4602-4838, 4605-5263, 4629-5261, 4630-4862, 4636-4889, 4650-5240, 4675-5269, 4678-4968, 4687-4961, 4687-4974, 4687-4991, 4687-4998, 4687-4999, 4689-4987, 4735-5270, 4740-5265, 4767-5265, 4791-5251,
	4822-5194, 4822-5250, 4835-5111, 4847-5254, 4871-5257, 4872-5251, 4872-5364, 4873-5511, 4907-5129, 4907-5241, 4907-5265, 4923-5191, 4923-5250, 4956-5166, 4985-5251, 5003-5214, 5003-5245, 5003-5321, 5009-5256

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
21/7472647CB1/7142	1-273, 54-343, 56-331, 72-379, 72-794, 81-307, 81-391, 81-459, 81-480, 81-486, 81-533, 81-569, 81-619, 83-633, 85-643, 92-609, 98-486, 104-556, 105-714, 137-707, 212-589, 256-833, 261-957, 290-680, 312-911, 374-1032, 379-934, 441-857, 453-1089, 457-925, 506-1073, 565-1195, 567-1065, 589-1219, 615-1162, 615-1178, 615-1201, 625-1175, 628-1060, 638-1213, 649-1226, 653-1269, 654-1226, 659-1282, 663-1076, 683-1232, 724-1017, 724-1246, 724-1306, 724-1311, 724-1314, 725-1387, 725-1417, 725-1476, 725-1528, 725-1543, 731-1345, 801-1256, 831-1424, 850-1422, 854-1417, 876-1332, 880-1422, 893-1427, 902-1508, 919-1490, 935-1415, 935-1591, 944-1552, 947-1508, 972-1539, 982-1552, 999-1687, 1017-1724, 1020-1552, 1034-1552, 1035-1552, 1037-1667, 1044-1552, 1052-1733, 1053-1564, 1057-1721, 1100-1552, 1108-1437, 1109-1386, 1125-1676, 1129-1552, 1146-1552, 1149-1422, 1149-1687, 1186-1799, 1199-1552, 1214-1760, 1214-1819, 1216-1552, 1217-1552, 1245-1314, 1248-1977, 1250-1552, 1281-1934, 1319-1552, 1322-1552, 1333-1925, 1336-1862, 1365-1866, 1390-1897, 1406-2003,
	1409-1977, 1412-1977, 1415-2008, 1427-2008, 1441-2008, 1452-2008, 1458-2005, 1527-2004, 1530-2008, 1558-2008, 1602-2008, 1628-1892, 1628-2008, 1641-2008, 1643-2008, 1649-2008, 1685-2008, 1694-2008, 1707-2553, 1731-2008, 1738-2008, 1746-2008, 1763-2008, 1810-2008, 1811-2008, 1819-2008, 1820-2008, 1826-2008, 1835-2008, 1849-2008, 1854-2008, 1862-2008, 1869-2008, 1876-2008, 1881-2008, 1900-2008, 1911-2008, 1924-2008, 2047-2551, 2056-2590, 2238-2950, 2364-2950, 2384-2950, 2668-3262, 3064-3345, 3286-3579, 3439-4034, 3543-3702, 3546-3705, 3706-4308, 3836-4495, 3959-4255, 4141-4729, 4221-4853, 4308-4566, 4308-4593, 4308-4915, 4407-5014, 4555-5162, 4865-5496, 4922-5554, 4986-5592, 5098-5624, 5229-5570, 5270-5544, 5270-5818, 5321-5953, 5347-5508, 5597-5867, 5597-6239, 5599-5871, 5702-6283, 5752-6015, 5752-6311, 5851-6117, 5903-6173, 5963-6216, 5963-6501, 5965-6488, 5984-6244, 6004-6250, 6020-6493, 6066-6091, 6085-6364, 6102-6291, 6105-6493, 6123-6501, 6132-6406, 6185-6428, 6216-6507, 6341-6598, 6425-6945, 6448-7128, 6505-6745, 6505-6782,
	6505-6783, 6524-7132, 6533-6825, 6592-6794, 6592-7120, 6601-7131, 6613-6856, 6613-7133, 6613-7142, 6679-6948, 6716-6977, 6730-6987

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
22/7472654CB1/6565	1-360, 1-372, 198-1217, 563-943, 715-1027, 1157-1292, 1157-1378, 1174-1217, 1174-1378, 1218-1323, 1324-1612, 1568-2264, 1568-2292, 1569-2318, 1569-2319, 1569-2331, 1569-2370, 1875-2438, 1940-2381, 2290-2593, 2324-2952, 2330-2952, 2331-2952, 2349-2952, 2361-2952, 2382-2684, 2475-2952, 2638-2947, 2684-3220, 2685-2814, 2742-3489, 2815-3019, 3015-3564, 3016-3289, 3016-3439, 3016-3558, 3016-3563, 3016-3564, 3016-3609, 3016-3684, 3018-3645, 3080-3579, 3104-3463, 3312-3968, 3312-3995, 3336-3844, 3387-3637, 3659-4388, 3686-3960, 3753-4298, 3773-4429, 3773-4478, 3797-4486, 3885-4453, 3885-4546, 3891-4508, 3981-4674, 4005-4551, 4041-4642, 4048-4724, 4072-4696, 4131-4563, 4140-4566, 4142-4718, 4153-4538, 4181-4843, 4182-4736, 4206-4484, 4206-4760, 4236-4795, 4242-4728, 4249-4793, 4251-4435, 4251-4837, 4256-4766, 4259-4824, 4277-4704, 4278-4743, 4286-4625, 4322-4963, 4399-4683, 4399-4915, 4405-4680, 4417-5127, 4489-5181, 4491-5127, 4528-4960, 4592-5023, 4593-5223, 4658-4914, 4674-4964, 4801-5467, 4802-5456, 5047-5601, 5067-5594,
	5078-5673, 5088-5525, 5187-5632, 5384-5965, 5434-6026, 5524-6100, 5576-6227, 5577-5814, 5578-5812, 5619-6251, 5622-5925, 5622-6137, 5636-6319, 5661-5896, 5695-5840, 5758-6200, 5765-6084, 5831-6539, 5833-6189, 5833-6212, 5833-6386, 5834-6232, 5941-6476, 5943-6547, 5969-6549, 6091-6565, 6295-6538
23/7480224CB1/1130	1-434, 1-436, 2-436, 144-794, 359-421, 359-426, 360-794, 645-1037, 795-1130
24/7481056CB1/2372	1-452, 8-181, 11-158, 11-184, 11-298, 12-431, 12-452, 14-452, 86-452, 140-428, 140-431, 193-452, 297-431, 364-1134, 404-431, 666-832, 700-1290, 1044-1797, 1046-1384, 1046-1398, 1046-1474, 1046-1507, 1046-1511, 1046-1526, 1046-1554, 1046-1558, 1046-1562, 1046-1576, 1046-1593, 1046-1618, 1046-1623, 1046-1635, 1046-1651, 1046-1657, 1046-1663, 1046-1683, 1046-1684, 1046-1711, 1046-1750, 1046-1774, 1046-1833, 1047-1816, 1048-1717, 1078-1158, 1087-1152, 1088-1683, 1124-1553, 1133-2351, 1174-1595, 1211-1979, 1231-1280, 1252-1748, 1307-2084, 1314-1787, 1371-1942, 1423-2299, 1436-2282, 1513-2165, 1564-2281, 1630-2159, 1862-2372, 1972-2349, 2252-2372

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
25/3750264CB1/4253	1-136, 1-578, 1-609, 188-608, 194-608, 494-809, 494-812, 494-813, 494-941, 494-973, 494-986, 494-1073, 494-1159, 494-1183, 494-1186, 494-1220, 497-812, 505-1226, 505-1250, 516-813, 541-813, 548-813, 558-813, 565-1124, 596-813, 609-812, 609-813, 609-1034, 609-1187, 609-1258, 609-1262, 612-1157, 613-1318, 633-813, 678-1266, 681-813, 691-813, 693-813, 694-813, 713-1456, 775-1380, 786-4102, 796-1375, 842-1439, 1081-1743, 1193-1459, 1193-1627, 1324-1745, 1380-1745, 1393-1745, 1460-1745, 1547-1735, 1547-1740, 1547-1743, 1547-1745, 1598-1994, 1610-1897, 1648-1897, 1658-2063, 1659-1791, 1752-2048, 1752-2170, 1788-2186, 1898-2044, 1898-2343, 2187-2478, 2187-2480, 2187-2605, 2187-2607, 2194-2527, 2194-2608, 2194-2674, 2194-2693, 2194-2771, 2194-2775, 2194-2780, 2194-2802, 2194-2803, 2194-2842, 2194-2847, 2194-2851, 2194-2856, 2194-2863, 2194-2874, 2194-2877, 2194-2879, 2194-2881, 2202-2888, 2205-2853, 2205-2944, 2210-2922, 2216-2929, 2216-2937, 2228-2816, 2295-2376, 2295-2404, 2295-2429, 2295-2433, 2295-2435, 2295-2464, 2295-2490, 2295-2492,
	2295-2498, 2295-2504, 2321-2983, 2326-3036, 2330-2909, 2356-2615, 2372-3025, 2390-3077, 2404-3116, 2407-2961, 2417-3148, 2432-2707, 2440-3230, 2452-3090, 2458-3174, 2469-3121, 2476-3116, 2479-2741, 2479-2986, 2489-3201, 2519-2998, 2524-3077, 2548-2662, 2560-3199, 2562-2785, 2578-3307, 2581-3108, 2607-3071, 2607-3141, 2608-2914, 2608-3163, 2608-3178, 2608-3190, 2608-3211, 2609-3166, 2609-3167, 2609-3178, 2609-3247, 2613-3292, 2617-2682, 2620-3166, 2622-2961, 2622-3197, 2623-3202, 2623-3209, 2625-3236, 2636-3267, 2638-3387, 2665-3385, 2677-3134, 2683-3191, 2703-3378, 2713-3491, 2721-3240, 2725-3395, 2752-3270, 2752-3414, 2793-3420, 2805-3069, 2805-3248, 2805-3409, 2828-3270, 2876-3574, 2890-3529, 2909-3064, 2909-3399, 2918-3404, 2923-3468, 2924-3416, 2928-3670, 2929-3632, 2948-3632, 2951-3518, 2952-3606, 2953-3390, 2961-3581, 2970-3632, 2974-3167, 2982-3728, 2991-3728, 2998-3620, 3006-3153, 3009-3336, 3016-3728, 3028-3541, 3031-3575, 3050-3697, 3061-3728, 3091-3474, 3095-3728, 3102-3728, 3107-3572, 3118-3572, 3125-3728, 3151-3850,
	3159-3743, 3172-3850, 3177-3850, 3181-3850, 3183-3850, 3194-3575, 3205-3850, 3220-3485, 3226-3850, 3243-3849, 3253-3850, 3255-3850, 3261-3850, 3262-3850, 3268-3849, 3276-3743, 3292-3850, 3306-3850, 3338-3850, 3342-3850, 3349-3806, 3360-3819, 3367-3831, 3377-3629, 3395-3850, 3404-3831, 3423-3850, 3426-3535, 3465-3849, 3487-3849, 3490-3849, 3507-3748, 3525-3849, 3529-3849, 3532-3655, 3687-3848, 3708-3849, 3727-3850, 3746-3834, 3746-3850, 3789-3840, 3842-4097, 3842-4174, 3842-4177, 3842-4253, 3846-4253, 3850-4253, 3851-4250, 3860-4253, 3883-4253, 3896-4253, 4038-4253, 4043-4253

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
26/1749735CB1/2681	1-608, 306-892, 416-561, 652-908, 652-1127, 652-1437, 653-1108, 716-1598, 847-1106, 1091-1684, 1160-1827, 1216-1791, 1222-1664, 1232-1855, 1297-1800, 1297-1931, 1303-1968, 1344-1934, 1361-1895, 1395-2061, 1559-2174, 1656-2347, 1871-2430, 2057-2681, 2093-2681, 2118-2681, 2124-2681, 2148-2681, 2211-2681
27/7473634CB1/4506	1-413, 206-743, 206-820, 206-872, 206-912, 414-604, 528-604, 594-1427, 594-1430, 605-692, 660-1430, 693-817, 814-1425, 818-939, 920-1430, 940-1156, 1157-2377, 1297-1844, 1297-2025, 1297-2037, 1871-2570, 1871-2579, 1871-2582, 1871-2611, 1871-2626, 2054-2927, 2158-2927, 2163-2927, 2337-2511, 2385-3194, 2402-3194, 2449-3194, 2475-3194, 2506-3194, 2727-3344, 2727-3377, 2732-3341, 2734-3547, 2900-3069, 3173-3630, 3227-3545, 3286-3634, 3430-3634, 3438-3635, 3457-3629, 3457-3633, 3457-3634, 3457-3635, 3486-4198, 3489-3664, 3489-4232, 3489-4242, 3489-4336, 3489-4506, 3490-3910
28/4767844CB1/1125	1-143, 1-153, 1-397, 1-708, 50-260, 50-474, 230-855, 243-759, 560-1125, 603-974, 612-1124, 726-992, 726-1013
29/7487584CB1/3062	1-273, 54-343, 56-331, 72-379, 72-794, 81-307, 81-391, 81-459, 81-480, 81-486, 81-533, 81-569, 81-619, 83-633, 85-643, 92-534, 92-609, 98-486, 104-556, 105-714, 137-707, 212-589, 256-833, 261-957, 290-680, 312-911, 358-575, 374-1032, 379-934, 393-606, 441-855, 441-857, 453-1089, 457-925, 489-606, 506-1073, 565-1195, 567-1065, 589-1219, 615-1162, 615-1171, 615-1178, 615-1201, 615-1212, 625-1175, 628-1060, 638-1213, 649-1226, 653-1269, 654-1226, 659-1282, 663-1076, 683-1232, 724-1017, 724-1246, 724-1306, 724-1311, 724-1314, 724-1387, 724-1417, 724-1476, 724-1543, 724-1564, 731-1345, 789-980, 789-1053, 801-1256, 831-1426, 850-1422, 854-1417, 859-1422, 875-1422, 876-1139, 876-1333, 882-1422, 889-1460, 891-1424, 892-1427, 902-1508, 919-1490, 935-1406, 935-1415, 935-1591, 936-1490, 944-1667, 947-1508, 972-1539, 982-1558, 999-1687, 1017-1724, 1020-1575, 1034-1575, 1035-1575, 1037-1667, 1044-1575, 1053-1564, 1057-1721, 1100-1575, 1108-1437, 1109-1386, 1116-1575, 1125-1676, 1129-1575, 1146-1576, 1149-1422, 1149-1687, 1186-1799, 1199-1575, 1214-1760,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1214-1819, 1216-1575, 1217-1575, 1248-1977, 1250-1575, 1281-1934, 1297-1575, 1319-1575, 1322-1575, 1333-1925, 1336-1862, 1365-1866, 1390-1897, 1406-2003, 1409-1977, 1412-1977, 1415-2163, 1426-1708, 1427-2112, 1440-2053, 1450-1657, 1452-2055, 1453-2143, 1454-1770, 1527-2179, 1530-2124, 1558-2086, 1601-2170, 1628-1892, 1628-2008, 1640-2096, 1643-2096, 1648-2401, 1685-2084, 1694-2228, 1727-2420, 1730-2280, 1746-2204, 1763-2287, 1809-2464, 1810-2449, 1811-2375, 1818-2291, 1820-2390, 1825-2309, 1830-2244, 1834-2425, 1846-2446, 1849-2449, 1850-1874, 1854-2487, 1859-1979, 1862-2465, 1869-2173, 1869-2441, 1876-2414, 1881-2449, 1884-2357, 1900-2492, 1911-2410, 1918-2138, 1922-2376, 1950-2700, 1959-2503, 2031-2602, 2045-2409, 2049-2323, 2053-2621, 2070-2655, 2070-2657, 2071-2459, 2079-2559, 2085-2575, 2085-2642, 2085-2643, 2167-2764, 2214-2621, 2214-2711, 2214-2712, 2217-2905, 2237-2779, 2238-3062, 2250-2776, 2253-2710, 2253-2760, 2253-2761, 2253-2764, 2253-2791, 2253-2805, 2253-2838, 2258-2764, 2261-2806, 2271-2796, 2310-2864, 2343-2938,
	2385-2893, 2385-2972, 2385-2973, 2394-2895, 2397-2806, 2427-2843, 2433-2792, 2433-3060, 2436-2806, 2445-2743, 2461-3046, 2605-2931, 2608-3010, 2667-3062
30/1468733CB1/1908	1-518, 10-507, 10-510, 10-511, 10-520, 10-531, 10-532, 10-537, 10-546, 10-559, 10-588, 14-749, 18-631, 19-520, 19-521, 19-522, 19-537, 19-550, 19-552, 19-581, 19-586, 19-613, 19-631, 19-663, 19-673, 21-581, 22-646, 26-591, 27-559, 30-641, 53-597, 60-604, 72-631, 78-541, 78-660, 78-742, 90-646, 92-636, 95-520, 98-641, 107-729, 114-729, 119-624, 123-748, 130-657, 141-712, 144-621, 150-749, 152-566, 152-717, 153-582, 154-634, 155-549, 158-744, 163-570, 165-749, 173-578, 174-683, 178-657, 182-537, 186-657, 187-677, 198-657, 214-269, 214-657, 232-657, 239-657, 239-749, 240-506, 241-749, 242-500, 242-501, 244-500, 248-690, 249-535, 254-737, 256-519, 256-604, 258-515, 258-537, 258-540, 266-555, 266-638, 266-744, 267-525, 268-529, 268-597, 270-597, 272-533, 273-749, 280-507, 280-552, 280-553, 280-749, 284-657, 292-737, 292-749, 294-641, 295-536, 295-576, 297-657, 303-749, 305-539, 305-552, 305-556, 305-573, 305-585, 305-594, 305-749, 316-601, 318-537, 321-749, 322-547, 323-749, 325-749, 328-749, 332-657, 334-657, 337-749, 340-595, 342-611, 347-749, 351-749, 354-741, 359-393,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
30	359-888, 360-749, 361-749, 364-652, 364-749, 369-749, 370-749, 371-637, 372-749, 374-597, 374-749, 376-658, 382-640, 390-749, 393-641, 398-657, 398-749, 399-687, 400-669, 400-682, 401-653, 401-657, 401-687, 403-744, 403-749, 409-749, 411-650, 411-749, 415-749, 416-668, 416-700, 418-664, 419-660, 422-637, 423-670, 423-724, 423-749, 436-748, 438-689, 438-744, 438-749, 457-713, 462-708, 463-749, 464-738, 465-657, 465-740, 465-742, 470-657, 470-733, 470-741, 473-696, 473-749, 479-749, 482-749, 488-726, 488-749, 490-742, 496-749, 501-749, 506-749, 508-734, 516-657, 523-597, 527-749, 528-747, 528-749, 534-749, 536-749, 538-561, 538-571, 538-576, 538-577, 538-578, 538-580, 538-581, 538-586, 538-590, 538-592, 538-593, 538-594, 538-595, 539-586, 539-591, 539-595, 540-574, 542-571, 550-749, 555-749, 597-746, 597-749, 598-619, 598-626, 598-630, 598-633, 598-636, 598-638, 598-641, 598-645, 598-646, 598-653, 598-654, 598-655, 598-687, 598-736, 598-741, 599-641, 599-651, 599-655, 599-687, 600-655, 608-655, 610-655, 615-655, 688-746, 688-749, 753-1262, 756-1171,
	783-1359, 784-1459, 806-1372, 813-1419, 822-868, 841-1515, 854-1442, 855-1431, 857-1433, 860-1453, 861-1405, 867-1428, 874-1446, 874-1472, 877-1544, 881-1436, 884-1759, 887-952, 887-1165, 887-1316, 887-1363, 887-1407, 888-1384, 889-1460, 896-1384, 897-1469, 898-953, 898-1481, 906-1371, 908-1469, 912-1759, 916-1357, 916-1398, 916-1406, 916-1423, 916-1460, 916-1490, 916-1514, 916-1517, 916-1526, 916-1527, 916-1535, 916-1580, 916-1590, 917-1509, 917-1534, 918-1513, 918-1526, 919-1509, 925-1583, 927-1534, 927-1587, 930-1387, 937-1480, 943-1414, 944-1589, 947-1525, 950-1427, 950-1578, 951-1587, 961-1495, 961-1590, 973-1519, 981-1473, 988-1488, 995-1535, 999-1601, 1004-1527, 1005-1606, 1006-1684, 1008-1406, 1010-1376, 1010-1531, 1013-1719, 1014-1500, 1014-1510, 1015-1615, 1020-1522, 1023-1550, 1030-1492, 1036-1594, 1038-1356, 1039-1569, 1042-1419, 1044-1494, 1046-1887, 1048-1537, 1049-1568, 1049-1594, 1053-1625, 1055-1364, 1057-1510, 1062-1662, 1064-1538, 1078-1360, 1080-1541, 1080-1630, 1080-1706, 1083-1658, 1084-1908,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1086-1367, 1091-1686, 1091-1733, 1092-1386, 1092-1742, 1094-1366, 1094-1434, 1095-1639, 1096-1368, 1096-1370, 1096-1374, 1096-1406, 1097-1289, 1097-1353, 1097-1409, 1097-1507, 1097-1571, 1097-1887, 1097-1895, 1097-1900, 1098-1376, 1098-1709, 1104-1408, 1105-1388, 1105-1429, 1111-1380, 1111-1488, 1112-1393, 1114-1524, 1116-1551, 1119-1512, 1119-1574, 1120-1401, 1122-1367, 1122-1372, 1122-1408, 1122-1433, 1123-1675, 1126-1444, 1128-1357, 1128-1396, 1128-1417, 1129-1378, 1129-1389, 1129-1466, 1129-1493, 1131-1381, 1133-1364, 1133-1542, 1133-1642, 1133-1742, 1136-1385, 1139-1354, 1141-1376, 1141-1452, 1141-1654, 1141-1861, 1147-1737, 1150-1399, 1151-1389, 1151-1395, 1151-1418, 1151-1423, 1154-1363, 1155-1450, 1155-1786, 1156-1780, 1158-1753, 1158-1801, 1160-1419, 1163-1426, 1163-1708, 1167-1442, 1167-1705, 1168-1371, 1168-1450, 1169-1410, 1169-1430, 1172-1685, 1173-1465, 1177-1401, 1179-1465, 1179-1484, 1180-1636, 1183-1418, 1184-1673, 1185-1509, 1186-1429, 1186-1589, 1187-1406, 1187-1412, 1187-1484, 1187-1584, 1187-1651, 1189-1409,
	1194-1449, 1194-1488, 1194-1795, 1196-1414, 1196-1445, 1196-1770, 1197-1480, 1202-1459, 1202-1461, 1202-1483, 1202-1494, 1202-1503, 1205-1426, 1205-1458, 1205-1462, 1206-1465, 1208-1861, 1211-1614, 1211-1833, 1213-1555, 1213-1897, 1214-1448, 1214-1759, 1216-1453, 1216-1474, 1217-1485, 1217-1515, 1218-1492, 1221-1465, 1221-1471, 1221-1801, 1223-1483, 1223-1489, 1223-1789, 1224-1505, 1224-1526, 1225-1700, 1226-1500, 1226-1502, 1226-1512, 1227-1571, 1228-1489, 1228-1503, 1228-1805, 1234-1494, 1234-1516, 1234-1517, 1234-1521, 1235-1479, 1235-1488, 1236-1506, 1324-1866, 1490-1531, 1663-1776
31/1652084CB1/1917	1-1386, 235-330, 235-419, 238-378, 438-493, 806-929, 828-983, 828-1359, 841-1619, 993-1243, 993-1661, 1111-1805, 1333-1582, 1333-1591, 1333-1709, 1333-1827, 1335-1837, 1343-1917, 1507-1861, 1536-1861
32/3456896CB1/1936	1-97, 1-290, 40-502, 70-699, 260-817, 304-936, 351-480, 351-675, 351-777, 351-904, 351-947, 351-964, 351-967, 351-977, 351-979, 351-982, 351-995, 351-1020, 351-1023, 351-1029, 351-1035, 351-1037, 351-1052, 351-1067, 351-1089, 357-986, 364-1105, 464-1097, 464-1118, 465-1163, 467-1096, 546-1296, 556-1182, 581-1299, 649-1329, 650-1299, 669-1093, 770-1006, 770-1089, 770-1116, 770-1160, 770-1170, 770-1227, 770-1304, 770-1327, 770-1332, 773-1456, 783-1427, 834-1579, 892-1032, 920-1394, 925-1513, 935-1413, 1057-1652, 1071-1777, 1072-1579, 1079-1665, 1094-1582, 1100-1608, 1123-1376, 1123-1564, 1127-1334, 1140-1920, 1190-1645, 1207-1754, 1207-1886, 1237-1570, 1257-1768, 1280-1552, 1280-1623, 1283-1771, 1301-1779, 1311-1922, 1311-1936, 1331-1936, 1335-1936, 1388-1936

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
17	7482256CB1	EOSINOT02
18	71973513CB1	OVARTUT02
19	7648238CB1	KIDNNOC01
20	1719204CB1	FIBPFEN06
21	7472647CB1	NERDTDN03
22	7472654CB1	FIBAUNT01
25	3750264CB1	SINTFER02
26	1749735CB1	BRATDIC01
27	7473634CB1	BRAUNOR01
28	4767844CB1	BRATNOT02
29	7487584CB1	BONEUNR01
30	1468733CB1	BRACNOK02
31	1652084CB1	PROSNOT16
32	3456896CB1	UTRSTUE01

Table 6

Library	Vector	Library Description
BONEUNR01	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from an untreated MG-63 cell line derived from an osteosarcoma tumor removed from a 14-year-old Caucasian male (donor A) and using mRNA isolated from sacral bone tumor tissue removed from an 18-year-old Caucasian female (donor B) during an exploratory laparotomy and soft tissue excision. Pathology indicated giant cell tumor of the sacrum in donor B. Donor B's history included pelvic joint pain, constipation, urinary incontinence, unspecified abdominal/pelvic symptoms, and a pelvic soft tissue malignant neoplasm. Family history included prostate cancer in donor B.
BRACNOK02	PSPORT1	This amplified and normalized library was constructed using RNA isolated from posterior cingulate tissue removed from an 85-year-old Caucasian female who died from myocardial infarction and retroperitoneal hemorrhage. Pathology indicated atherosclerosis, moderate to severe, involving the circle of Willis, middle cerebral, basilar and vertebral arteries; infarction, remote, left dentate nucleus; and amyloid plaque deposition consistent with age. There was mild to moderate leptomeningeal fibrosis, especially over the convexity of the frontal lobe. There was mild generalized atrophy involving all lobes. The white matter was mildly thinned. Cortical thickness in the temporal lobes, both maximal and minimal, was slightly reduced. The substantia nigra pars compacta appeared mildly depigmented. Patient history included COPD, hypertension, and recurrent deep venous thrombosis. 6.4 million independent clones from this amplified library were normalized in one round using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791.
BRATDIC01	pINCY	This large size-fractionated library was constructed using RNA isolated from diseased brain tissue removed from the left temporal lobe of a 27-year-old Caucasian male during a brain lobectomy. Pathology for the left temporal lobe, including the mesial temporal structures, indicated focal, marked pyramidal cell loss and gliosis in hippocampal sector CA1, consistent with mesial temporal sclerosis. The left frontal lobe showed a focal deep white matter lesion, characterized by marked gliosis, calcifications, and hemosiderin-laden macrophages, consistent with a remote perinatal injury. The frontal lobe tissue also showed mild to moderate generalized gliosis, predominantly subpial and subcortical, consistent with chronic seizure disorder. GFAP was positive for astrocytes. The patient presented with intractable epilepsy, focal epilepsy, hemiplegia, and an unspecified brain injury. Patient history included cerebral palsy, abnormality of gait, depressive disorder, and tobacco abuse in remission. Previous surgeries included tendon transfer. Patient medications included minocycline hydrochloride, Tegretol, phenobarbital, vitamin C, Pepcid, and Pevaryl. Family history included brain cancer i
BRATNOT02	pINCY	Library was constructed using RNA isolated from superior temporal cortex tissue removed from the brain of a 35-year-old Caucasian male. No neuropathology was found. Patient history included dilated cardiomyopathy, congestive heart failure, and an enlarged spleen and liver.

Table 6

Library	Vector	Library Description
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive
EOSINOT02	PSPORT	Library was constructed using RNA isolated from pooled eosinophils obtained from allergic asthmatic individuals.
FIBAUNT01	pINCY	Library was constructed using RNA isolated from untreated aortic adventitial fibroblasts obtained from a 48- year-old Caucasian male.
FIBPFEN06	pINCY	The normalized prostate stromal fibroblast tissue libraries were constructed from 1.56 million independent clones from a prostate fibroblast library. Starting RNA was made from fibroblasts of prostate stroma removed from a male fetus, who died after 26 weeks' gestation. The libraries were normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48-hours/round)reannealing hybridization was used. The library was then linearized and recircularized to select for insert containing clones as follows: plasmid DNA was prepped from approximately 1 million clones from the normalized prostate stromal fibroblast tissue libraries following soft agar transformation.
KIDNNOC01	pINCY	This large size-fractionated library was constructed using RNA isolated from pooled left and right kidney tissue removed from a Caucasian male fetus, who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation.

Table 6

Library	Vector	Library Description
NERDTDN03	pINCY	This normalized dorsal root ganglion tissue library was constructed from 1.05 million independent clones from a dorsal root ganglion tissue library. Starting RNA was made from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema, acute bronchopneumonia, bilateral pleural effusions, pericardial effusion, and malignant lymphoma (natural killer cell type). The patient presented with pyrexia of unknown origin, malaise, fatigue, and gastrointestinal bleeding. Patient history included probable cytomegalovirus infection, liver congestion, and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, respiratory failure, pneumonia of the left lung, natural killer cell lymphoma of the pharynx, Bell's palsy, and tobacco and alcohol abuse. Previous surgeries included colonoscopy, closed colon biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy. Patient medications included Diflucan (fluconazole), Deltasone (prednisone), hydrocodone, Lortab, Alprazolam, Reazodone, ProMace-Cytabom, Etoposide, Cisplatin, Cytarabine, and dexamethasone. The patient received radiation therapy and multip
OVARTUT02	pINCY	Library was constructed using RNA isolated from ovarian tumor tissue removed from a 51-year-old Caucasian female during an exploratory laparotomy, total abdominal hysterectomy, salpingo-oophorectomy, and an incidental appendectomy. Pathology indicated mucinous cystadenoma presenting as a multiloculated neoplasm involving the entire left ovary. The right ovary contained a follicular cyst and a hemorrhagic corpus luteum. The uterus showed proliferative endometrium and a single intramural leiomyoma. The peritoneal biopsy indicated benign glandular inclusions consistent with endosalpingiosis. Family history included atherosclerotic coronary artery disease, benign hypertension, breast cancer, and uterine cancer.
PROSNOT16	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis, and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
SINTFER02	pINCY	This random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from fetal demise.

Table 6

Library	Vector	Library Description
UTRSTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from uterus tumor tissue removed a 37-year-old Black female during myomectomy, dilation and curettage, right fimbrial region biopsy, and incidental appendectomy. Pathology indicated multiple (12) uterine leiomyomata. A fimbrial cyst was identified. The patient presented with deficiency anemia, an umbilical hernia, and premenopausal menorrhagia. Patient history included premenopausal menorrhagia and sarcoidosis of the lung. Previous surgeries included hysteroscopy, dilation and curettage, and an endoscopic lung biopsy. Patient medications included Chromagen and Claritin. Family history included acute myocardial infarction and atherosclerotic coronary artery disease in the father.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less; Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6; Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less; Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less; Signal peptide hits: Score= 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain

reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

19. A method for treating a disease or condition associated with decreased expression of functional PMMM, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional PMMM, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional

PMMM, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target

polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of PMMM in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of PMMM in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of PMMM in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of

claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID

NO:32.

<110> INCYTE GENOMICS, INC.
Warren, Bridget A.
Honchell,Cynthia D.
Lu,Yan
Walia, Narinder K.
Burford, Neil
Delegeane,Angelo M.
Gandhi,Ameena R.
Baughn, Mariah R.
Griffin,Jennifer A.
Gietzen, Kimberly J.
Lu,Dyng Aina M.
Ison,Craig H.
Ramkumar,Jayalaxmi
Tang, Y. Tom
Lal, Preeti G.
Borowsky, Mark L.
Duggan, Brendan M.
Hafalia, April J.A.
Arvizu,Chandra
Thangavelu,Kavitha
Yao, Monique G.
Elliott, Vicki S.
Ding, Li
Yue, Henry
Lee, Sally
Swarnakar, Anita
Tran, Uyen K.
Xu, Yuming

<120> PROTEIN MODIFICATION AND MAINTENANCE MOLECULES

<130> PI-0361 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/265,705
60/266,762
60/269,581
60/271,198
60/272,813
60/275,586
60/278,508
60/280,539

<151> 2001-01-31
2001-02-05
2001-02-16
2001-02-23
2001-03-01
2001-03-13
2001-03-23
2001-03-30

<160> 32
<170> PERL Program

<210> 1
<211> 269
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7482256CD1

<400> 1
Met Gly Ala Arg Gly Ala Leu Leu Leu Ala Leu Leu Leu Ala Arg
1 5 10 15
Ala Gly Leu Gly Lys Pro Glu Ala Cys Gly His Arg Glu Ile His
20 25 30
Ala Leu Val Ala Gly Gly Val Glu Ser Ala Arg Gly Arg Trp Pro
35 40 45
Trp Gln Ala Ser Leu Arg Leu Arg Arg Arg His Arg Cys Gly Gly
50 55 60
Ser Leu Leu Ser Arg Arg Trp Val Leu Ser Ala Ala His Cys Phe
65 70 75
Gln Asn Ser Arg Tyr Lys Val Gln Asp Ile Ile Val Asn Pro Asp
80 85 90
Ala Leu Gly Val Leu Arg Asn Asp Ile Ala Leu Leu Arg Leu Ala
95 100 105
Ser Ser Val Thr Tyr Asn Ala Tyr Ile Gln Pro Ile Cys Ile Glu
110 115 120
Ser Ser Thr Phe Asn Phe Val His Arg Pro Asp Cys Trp Val Thr
125 130 135
Gly Trp Gly Leu Ile Ser Pro Ser Gly Thr Pro Leu Pro Pro Pro
140 145 150
Tyr Asn Leu Arg Glu Ala Gln Val Thr Ile Leu Asn Asn Thr Arg
155 160 165
Cys Asn Tyr Leu Phe Glu Gln Pro Ser Ser Arg Ser Met Ile Trp
170 175 180
Asp Ser Met Phe Cys Ala Gly Ala Glu Asp Gly Ser Val Asp Thr
185 190 195
Cys Lys Gly Asp Ser Gly Gly Pro Leu Val Cys Asp Lys Asp Gly
200 205 210
Leu Trp Tyr Gln Val Gly Ile Val Ser Trp Gly Met Asp Cys Gly
215 220 225
Gln Pro Asn Arg Pro Gly Val Tyr Thr Asn Ile Ser Val Tyr Phe
230 235 240
His Trp Ile Arg Arg Val Met Ser His Ser Thr Pro Arg Pro Asn
245 250 255
Pro Pro Gln Leu Leu Leu Leu Leu Ala Leu Leu Trp Ala Pro
260 265

<210> 2
<211> 379
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71973513CD1

<400> 2
Met Arg Gly Leu Val Val Phe Leu Ala Val Phe Ala Leu Ser Glu
1 5 10 15
Val Asn Ala Ile Thr Arg Val Pro Leu His Lys Gly Lys Ser Leu
20 25 30
Arg Arg Ala Leu Lys Glu Arg Arg Leu Leu Glu Asp Phe Leu Arg
35 40 45
Asn His His Tyr Ala Val Ser Arg Lys His Ser Ser Ser Gly Val
50 55 60
Val Ala Ser Glu Ser Leu Thr Asn Tyr Leu Asp Cys Gln Tyr Phe
65 70 75
Gly Lys Ile Tyr Ile Gly Thr Leu Pro Gln Lys Phe Thr Leu Val

```
                80                  85                  90
Phe Asp Thr Gly Ser Pro Asp Ile Trp Val Pro Ser Val Tyr Cys
                95                 100                 105
Asn Ser Asp Ala Cys Gln Asn His Gln Arg Phe Asp Pro Ser Lys
               110                 115                 120
Ser Ser Thr Gln Asn Met Gly Lys Ser Leu Ser Ile Gln Tyr Gly
               125                 130                 135
Thr Gly Ser Met Arg Gly Leu Leu Gly Tyr Asp Thr Val Thr Val
               140                 145                 150
Ser Asn Ile Val Asp Pro His Gln Thr Val Gly Leu Ser Thr Gln
               155                 160                 165
Glu Pro Gly Asp Val Phe Thr Tyr Ser Glu Phe Asp Gly Ile Leu
               170                 175                 180
Gly Leu Ala Tyr Pro Ser Leu Ala Ser Glu Tyr Ala Leu Arg Leu
               185                 190                 195
Gly Phe Arg Asn Asp Gln Gly Ser Met Leu Thr Leu Arg Ala Ile
               200                 205                 210
Asp Leu Ser Tyr Tyr Thr Gly Ser Leu His Trp Ile Pro Met Thr
               215                 220                 225
Ala Arg Ile Leu Ala Val His Cys Gly Gln Glu Gly Pro Gly Glu
               230                 235                 240
Gly Gly Leu Asp Glu Ala Ile Leu His Thr Phe Gly Ser Val Ile
               245                 250                 255
Ile Asp Gly Val Val Val Ala Cys Asp Gly Gly Cys Gln Ala Ile
               260                 265                 270
Leu Asp Thr Gly Thr Ser Leu Leu Val Gly Pro Gly Gly Asn Ile
               275                 280                 285
Leu Asn Ile Gln Gln Ala Ile Gly Arg Thr Ala Gly Gln Tyr Asn
               290                 295                 300
Glu Phe Asp Ile Asp Cys Gly Arg Leu Ser Ser Ile Pro Thr Ala
               305                 310                 315
Val Phe Glu Ile His Gly Lys Lys Tyr Pro Leu Pro Pro Ser Ala
               320                 325                 330
Tyr Thr Ser Gln Asp Gln Gly Phe Cys Thr Ser Gly Phe Gln Gly
               335                 340                 345
Asp Tyr Ser Ser Gln Gln Trp Ile Leu Gly Asn Val Phe Ile Trp
               350                 355                 360
Glu Tyr Tyr Ser Val Phe Asp Arg Thr Asn Asn Arg Val Gly Leu
               365                 370                 375
Ala Lys Ala Val
```

<210> 3
<211> 398
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7648238CD1

<400> 3

```
Met Leu Ser Ser Pro Gly Val Ala Ala Ala Val Val Thr Ala Leu
  1               5                  10                  15
Glu Asp Val Phe Gln Ala Leu Gly Phe Glu Ser Cys Glu Arg Arg
               20                  25                  30
Glu Val Pro Val Gln Gly Phe Leu Glu Glu Leu Ala Trp Phe Gln
               35                  40                  45
Glu Gln Leu Asp Ala His Gly Arg Pro Val Gly Gly Gln Leu Arg
               50                  55                  60
Gln Pro Gln Gln Leu Val Arg Glu Leu Ser Gly Cys Arg Ala Leu
               65                  70                  75
Arg Gly Cys Pro Lys Val Phe Leu Leu Leu Ser Ser Gly Pro Gly
```

```
                80                  85                  90
Ser Ser Leu Glu Pro Gly Ala Phe Leu Ala Gly Leu Arg Glu Leu
                95                 100                 105
Cys Gly Arg Ser Pro His Trp Ser Leu Val Gln Leu Leu Thr Lys
               110                 115                 120
Leu Phe Arg Arg Val Ala Glu Glu Ser Ala Gly Gly Thr Cys Cys
               125                 130                 135
Pro Val Leu Arg Ser Ser Leu Arg Gly Ala Leu Cys Leu Gly Gly
               140                 145                 150
Val Glu Pro Trp Arg Pro Glu Pro Ala Pro Gly Pro Ser Thr Gln
               155                 160                 165
Tyr Asp Leu Ser Lys Ala Arg Ala Ala Leu Leu Leu Ala Val Ile
               170                 175                 180
Gln Gly Arg Pro Gly Ala Gln His Asp Val Glu Ala Leu Gly Gly
               185                 190                 195
Leu Cys Trp Ala Leu Gly Phe Glu Thr Thr Val Arg Thr Asp Pro
               200                 205                 210
Thr Ala Gln Ala Phe Gln Glu Glu Leu Ala Gln Phe Arg Glu Gln
               215                 220                 225
Leu Asp Thr Cys Arg Gly Pro Val Ser Cys Ala Leu Val Ala Leu
               230                 235                 240
Met Ala His Gly Gly Pro Arg Gly Gln Leu Leu Gly Ala Asp Gly
               245                 250                 255
Gln Glu Val Gln Pro Glu Ala Leu Met Gln Glu Leu Ser Arg Cys
               260                 265                 270
Gln Val Leu Gln Gly Arg Pro Lys Ile Phe Leu Leu Gln Ala Cys
               275                 280                 285
Arg Gly Gly Asn Arg Asp Ala Gly Val Gly Pro Thr Ala Leu Pro
               290                 295                 300
Trp Tyr Trp Ser Trp Leu Arg Ala Pro Pro Ser Val Pro Ser His
               305                 310                 315
Ala Asp Val Leu Gln Ile Tyr Ala Glu Ala Gln Gly Tyr Val Ala
               320                 325                 330
Tyr Arg Asp Asp Lys Gly Ser Asp Phe Ile Gln Thr Leu Val Glu
               335                 340                 345
Val Leu Arg Ala Asn Pro Gly Arg Asp Leu Leu Glu Leu Leu Thr
               350                 355                 360
Glu Val Asn Arg Arg Val Cys Glu Gln Glu Val Leu Gly Pro Asp
               365                 370                 375
Cys Asp Glu Leu Arg Lys Ala Cys Leu Glu Ile Arg Ser Ser Leu
               380                 385                 390
Arg Arg Arg Leu Cys Leu Gln Ala
               395
```

<210> 4
<211> 1221
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1719204CD1

```
<400> 4
Met Ala Pro Leu Arg Ala Leu Leu Ser Tyr Leu Leu Pro Leu His
  1               5                  10                  15
Cys Ala Leu Cys Ala Ala Ala Gly Ser Arg Thr Pro Glu Leu His
               20                  25                  30
Leu Ser Gly Lys Leu Ser Asp Tyr Gly Val Thr Val Pro Cys Ser
               35                  40                  45
Thr Asp Phe Arg Gly Arg Phe Leu Ser His Val Val Ser Gly Pro
               50                  55                  60
Ala Ala Ala Ser Ala Gly Ser Met Val Val Asp Thr Pro Pro Thr
```

```
                65                  70                  75
Leu Pro Arg His Ser Ser His Leu Arg Val Ala Arg Ser Pro Leu
                80                  85                  90
His Pro Gly Gly Thr Leu Trp Pro Gly Arg Val Gly Arg His Ser
                95                 100                 105
Leu Tyr Phe Asn Val Thr Val Phe Gly Lys Glu Leu His Leu Arg
               110                 115                 120
Leu Arg Pro Asn Arg Arg Leu Val Val Pro Gly Ser Ser Val Glu
               125                 130                 135
Trp Gln Glu Asp Phe Arg Glu Leu Phe Arg Gln Pro Leu Arg Gln
               140                 145                 150
Glu Cys Val Tyr Thr Gly Gly Val Thr Gly Met Pro Gly Ala Ala
               155                 160                 165
Val Ala Ile Ser Asn Cys Asp Gly Leu Ala Gly Leu Ile Arg Thr
               170                 175                 180
Asp Ser Thr Asp Phe Phe Ile Glu Pro Leu Glu Arg Gly Gln Gln
               185                 190                 195
Glu Lys Glu Ala Ser Gly Arg Thr His Val Val Tyr Arg Arg Glu
               200                 205                 210
Ala Val Gln Gln Glu Trp Ala Glu Pro Asp Gly Asp Leu His Asn
               215                 220                 225
Glu Ala Phe Gly Leu Gly Asp Leu Pro Asn Leu Leu Gly Leu Val
               230                 235                 240
Gly Asp Gln Leu Gly Asp Thr Glu Arg Lys Arg Arg His Ala Lys
               245                 250                 255
Pro Gly Ser Tyr Ser Ile Glu Val Leu Leu Val Val Asp Asp Ser
               260                 265                 270
Val Val Arg Phe His Gly Lys Glu His Val Gln Asn Tyr Val Leu
               275                 280                 285
Thr Leu Met Asn Ile Val Asp Glu Ile Tyr His Asp Glu Ser Leu
               290                 295                 300
Gly Val His Ile Asn Ile Ala Leu Val Arg Leu Ile Met Val Gly
               305                 310                 315
Tyr Arg Gln Ser Leu Ser Leu Ile Glu Arg Gly Asn Pro Ser Arg
               320                 325                 330
Ser Leu Glu Gln Val Cys Arg Trp Ala His Ser Gln Gln Arg Gln
               335                 340                 345
Asp Pro Ser His Ala Glu His His Asp His Val Val Phe Leu Thr
               350                 355                 360
Arg Gln Asp Phe Gly Pro Ser Gly Tyr Ala Pro Val Thr Gly Met
               365                 370                 375
Cys His Pro Leu Arg Ser Cys Ala Leu Asn His Glu Asp Gly Phe
               380                 385                 390
Ser Ser Ala Phe Val Ile Ala His Glu Thr Gly His Val Leu Gly
               395                 400                 405
Met Glu His Asp Gly Gln Gly Asn Gly Cys Ala Asp Glu Thr Ser
               410                 415                 420
Leu Gly Ser Val Met Ala Pro Leu Val Gln Ala Ala Phe His Arg
               425                 430                 435
Phe His Trp Ser Arg Cys Ser Lys Leu Glu Leu Ser Arg Tyr Leu
               440                 445                 450
Pro Ser Tyr Asp Cys Leu Leu Asp Asp Pro Phe Asp Pro Ala Trp
               455                 460                 465
Pro Gln Pro Pro Glu Leu Pro Gly Ile Asn Tyr Ser Met Asp Glu
               470                 475                 480
Gln Cys Arg Phe Asp Phe Gly Ser Gly Tyr Gln Thr Cys Leu Ala
               485                 490                 495
Phe Arg Thr Phe Glu Pro Cys Lys Gln Leu Trp Cys Ser His Pro
               500                 505                 510
Asp Asn Pro Tyr Phe Cys Lys Thr Lys Lys Gly Pro Pro Leu Asp
               515                 520                 525
Gly Thr Glu Cys Ala Pro Gly Lys Trp Cys Phe Lys Gly His Cys
               530                 535                 540
```

```
Ile Trp Lys Ser Pro Glu Gln Thr Tyr Gly Gln Asp Gly Gly Trp
545                 550                 555
Ser Ser Trp Thr Lys Phe Gly Ser Cys Ser Arg Ser Cys Gly Gly
560                 565                 570
Gly Val Arg Ser Arg Ser Arg Ser Cys Asn Asn Pro Ser Leu Trp
575                 580                 585
Ser Arg Pro Cys Leu Gly Pro Met Phe Glu Tyr Gln Val Cys Asn
590                 595                 600
Ser Glu Glu Cys Pro Gly Thr Tyr Glu Asp Phe Arg Ala Gln Gln
605                 610                 615
Cys Ala Lys Arg Asn Ser Tyr Tyr Val His Gln Asn Ala Lys His
620                 625                 630
Ser Trp Val Pro Tyr Glu Pro Asp Asp Asp Ala Gln Lys Cys Glu
635                 640                 645
Leu Ile Cys Gln Ser Ala Asp Thr Gly Asp Val Val Phe Met Asn
650                 655                 660
Gln Val Val His Asp Gly Thr Arg Cys Ser Tyr Arg Asp Pro Tyr
665                 670                 675
Ser Val Cys Ala Arg Gly Glu Cys Val Pro Val Gly Cys Asp Lys
680                 685                 690
Glu Val Gly Ser Met Lys Ala Asp Asp Lys Cys Gly Val Cys Gly
695                 700                 705
Gly Asp Asn Ser His Cys Arg Thr Val Lys Gly Thr Leu Gly Lys
710                 715                 720
Ala Ser Lys Gln Ala Gly Ala Leu Lys Leu Val Gln Ile Pro Ala
725                 730                 735
Gly Ala Arg His Ile Gln Ile Glu Ala Leu Glu Lys Ser Pro His
740                 745                 750
Arg Ser Val Val Lys Asn Gln Val Thr Gly Ser Phe Ile Leu Asn
755                 760                 765
Pro Lys Gly Lys Glu Ala Thr Ser Arg Thr Phe Thr Ala Met Gly
770                 775                 780
Leu Glu Trp Glu Asp Ala Val Glu Asp Ala Lys Glu Ser Leu Lys
785                 790                 795
Thr Ser Gly Pro Leu Pro Glu Ala Ile Ala Ile Leu Ala Leu Pro
800                 805                 810
Pro Thr Glu Gly Gly Pro Arg Ser Ser Leu Ala Tyr Lys Tyr Val
815                 820                 825
Ile His Glu Asp Leu Leu Pro Leu Ile Gly Ser Asn Asn Val Leu
830                 835                 840
Leu Glu Glu Met Asp Thr Tyr Glu Trp Ala Leu Lys Ser Trp Ala
845                 850                 855
Pro Cys Ser Lys Ala Cys Gly Gly Gly Ile Gln Phe Thr Lys Tyr
860                 865                 870
Gly Cys Arg Arg Arg Arg Asp His His Met Val Gln Arg His Leu
875                 880                 885
Cys Asp His Lys Lys Arg Pro Lys Pro Ile Arg Arg Arg Cys Asn
890                 895                 900
Gln His Pro Cys Ser Gln Pro Val Trp Val Thr Glu Glu Trp Gly
905                 910                 915
Ala Cys Ser Arg Ser Cys Gly Lys Leu Gly Val Gln Thr Arg Gly
920                 925                 930
Ile Gln Cys Leu Leu Pro Leu Ser Asn Gly Thr His Lys Val Met
935                 940                 945
Pro Ala Lys Ala Cys Ala Gly Asp Arg Pro Glu Ala Arg Arg Pro
950                 955                 960
Cys Leu Arg Val Pro Cys Pro Ala Gln Trp Arg Leu Gly Ala Trp
965                 970                 975
Ser Gln Cys Ser Ala Thr Cys Gly Glu Gly Ile Gln Gln Arg Gln
980                 985                 990
Val Val Cys Arg Thr Asn Ala Asn Ser Leu Gly His Cys Glu Gly
995                1000                1005
Asp Arg Pro Asp Thr Val Gln Val Cys Ser Leu Pro Ala Cys Gly
```

1010 1015 1020
Gly Asn His Gln Asn Ser Thr Val Arg Ala Asp Val Trp Glu Leu
1025 1030 1035
Gly Thr Pro Glu Gly Gln Trp Val Pro Gln Ser Glu Pro Leu His
1040 1045 1050
Pro Ile Asn Lys Ile Ser Ser Thr Glu Pro Cys Thr Gly Asp Arg
1055 1060 1065
Ser Val Phe Cys Gln Met Glu Val Leu Asp Arg Tyr Cys Ser Ile
1070 1075 1080
Pro Gly Tyr His Arg Leu Cys Cys Val Ser Cys Ile Lys Lys Ala
1085 1090 1095
Ser Gly Pro Asn Pro Gly Pro Asp Pro Gly Pro Thr Ser Leu Pro
1100 1105 1110
Pro Phe Ser Thr Pro Gly Ser Pro Leu Pro Gly Pro Gln Asp Pro
1115 1120 1125
Ala Asp Ala Ala Glu Pro Pro Gly Lys Pro Thr Gly Ser Glu Asp
1130 1135 1140
His Gln His Gly Arg Ala Thr Gln Leu Pro Gly Ala Leu Asp Thr
1145 1150 1155
Ser Ser Pro Gly Thr Gln His Pro Phe Ala Pro Glu Thr Pro Ile
1160 1165 1170
Pro Gly Ala Ser Trp Ser Ile Ser Pro Thr Thr Pro Gly Gly Leu
1175 1180 1185
Pro Trp Gly Trp Thr Gln Thr Pro Thr Pro Val Pro Glu Asp Lys
1190 1195 1200
Gly Gln Pro Gly Glu Asp Leu Arg His Pro Gly Thr Ser Leu Pro
1205 1210 1215
Ala Ala Ser Pro Val Thr
1220

<210> 5
<211> 1537
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472647CD1

<400> 5
Met Glu Cys Cys Arg Arg Ala Thr Pro Gly Thr Leu Leu Leu Phe
1 5 10 15
Leu Ala Phe Leu Leu Leu Ser Ser Arg Thr Ala Arg Ser Glu Glu
20 25 30
Asp Arg Asp Gly Leu Trp Asp Ala Trp Gly Pro Trp Ser Glu Cys
35 40 45
Ser Arg Thr Cys Gly Gly Gly Ala Ser Tyr Ser Leu Arg Arg Cys
50 55 60
Leu Ser Ser Lys Ser Cys Glu Gly Arg Asn Ile Arg Tyr Arg Thr
65 70 75
Cys Ser Asn Val Asp Cys Pro Pro Glu Ala Gly Asp Phe Arg Ala
80 85 90
Gln Gln Cys Ser Ala His Asn Asp Val Lys His His Gly Gln Phe
95 100 105
Tyr Glu Trp Leu Pro Val Ser Asn Asp Pro Asp Asn Pro Cys Ser
110 115 120
Leu Lys Cys Gln Ala Lys Gly Thr Thr Leu Val Val Glu Leu Ala
125 130 135
Pro Lys Val Leu Asp Gly Thr Arg Cys Tyr Thr Glu Ser Leu Asp
140 145 150
Met Cys Ile Ser Gly Leu Cys Gln Ile Val Gly Cys Asp His Gln
155 160 165
Leu Gly Ser Thr Val Lys Glu Asp Asn Cys Gly Val Cys Asn Gly

```
                170                 175                 180
Asp Gly Ser Thr Cys Arg Leu Val Arg Gly Gln Tyr Lys Ser Gln
                185                 190                 195
Leu Ser Ala Thr Lys Ser Asp Asp Thr Val Val Ala Ile Pro Tyr
                200                 205                 210
Gly Ser Arg His Ile Arg Leu Val Leu Lys Gly Pro Asp His Leu
                215                 220                 225
Tyr Leu Glu Thr Lys Thr Leu Gln Gly Thr Lys Gly Glu Asn Ser
                230                 235                 240
Leu Ser Ser Thr Gly Thr Phe Leu Val Asp Asn Ser Ser Val Asp
                245                 250                 255
Phe Gln Lys Phe Pro Asp Lys Glu Ile Leu Arg Met Ala Gly Pro
                260                 265                 270
Leu Thr Ala Asp Phe Ile Val Lys Ile Arg Asn Ser Gly Ser Ala
                275                 280                 285
Asp Ser Thr Val Gln Phe Ile Phe Tyr Gln Pro Ile Ile His Arg
                290                 295                 300
Trp Arg Glu Thr Asp Phe Phe Pro Cys Ser Ala Thr Cys Gly Gly
                305                 310                 315
Gly Tyr Gln Leu Thr Ser Ala Glu Cys Tyr Asp Leu Arg Ser Asn
                320                 325                 330
Arg Val Val Ala Asp Gln Tyr Cys His Tyr Tyr Pro Glu Asn Ile
                335                 340                 345
Lys Pro Lys Pro Lys Leu Gln Glu Cys Asn Leu Asp Pro Cys Pro
                350                 355                 360
Ala Ser Asp Gly Tyr Lys Gln Ile Met Pro Tyr Asp Leu Tyr His
                365                 370                 375
Pro Leu Pro Arg Trp Glu Ala Thr Pro Trp Thr Ala Cys Ser Ser
                380                 385                 390
Ser Cys Gly Gly Asp Ile Gln Ser Arg Ala Val Ser Cys Val Glu
                395                 400                 405
Glu Asp Ile Gln Gly His Val Thr Ser Val Glu Glu Trp Lys Cys
                410                 415                 420
Met Tyr Thr Pro Lys Met Pro Ile Ala Gln Pro Cys Asn Ile Phe
                425                 430                 435
Asp Cys Pro Lys Trp Leu Ala Gln Glu Trp Ser Pro Cys Thr Val
                440                 445                 450
Thr Cys Gly Gln Gly Leu Arg Tyr Arg Val Val Leu Cys Ile Asp
                455                 460                 465
His Arg Gly Met His Thr Gly Gly Cys Ser Pro Lys Thr Lys Pro
                470                 475                 480
His Ile Lys Glu Glu Cys Ile Val Pro Thr Pro Cys Tyr Lys Pro
                485                 490                 495
Lys Glu Lys Leu Pro Val Glu Ala Lys Leu Pro Trp Phe Lys Gln
                500                 505                 510
Ala Gln Glu Leu Glu Glu Gly Ala Ala Val Ser Glu Glu Pro Ser
                515                 520                 525
Phe Ile Pro Glu Ala Trp Ser Ala Cys Thr Val Thr Cys Gly Val
                530                 535                 540
Gly Thr Gln Val Arg Ile Val Arg Cys Gln Val Leu Leu Ser Phe
                545                 550                 555
Ser Gln Ser Val Ala Asp Leu Pro Ile Asp Glu Cys Glu Gly Pro
                560                 565                 570
Lys Pro Ala Ser Gln Arg Ala Cys Tyr Ala Gly Pro Cys Ser Gly
                575                 580                 585
Glu Ile Pro Glu Phe Asn Pro Asp Glu Thr Asp Gly Leu Phe Gly
                590                 595                 600
Gly Leu Gln Asp Phe Asp Glu Leu Tyr Asp Trp Glu Tyr Glu Gly
                605                 610                 615
Phe Thr Lys Cys Ser Glu Ser Cys Gly Gly Gly Pro Gly Arg Pro
                620                 625                 630
Ser Thr Lys His Ser Pro His Ile Ala Ala Ala Arg Lys Val Tyr
                635                 640                 645
```

```
Ile Gln Thr Arg Arg Gln Arg Lys Leu His Phe Val Val Gly Gly
                650                 655                 660
Phe Ala Tyr Leu Leu Pro Lys Thr Ala Val Val Leu Arg Cys Pro
                665                 670                 675
Ala Arg Arg Val Arg Lys Pro Leu Ile Thr Trp Glu Lys Asp Gly
                680                 685                 690
Gln His Leu Ile Ser Ser Thr His Val Thr Val Ala Pro Phe Gly
                695                 700                 705
Tyr Leu Lys Ile His Arg Leu Lys Pro Ser Asp Ala Gly Val Tyr
                710                 715                 720
Thr Cys Ser Ala Gly Pro Ala Arg Glu His Phe Val Ile Lys Leu
                725                 730                 735
Ile Gly Gly Asn Arg Lys Leu Val Ala Arg Pro Leu Ser Pro Arg
                740                 745                 750
Ser Glu Glu Glu Val Leu Ala Gly Arg Lys Gly Gly Pro Lys Glu
                755                 760                 765
Ala Leu Gln Thr His Lys His Gln Asn Gly Ile Phe Ser Asn Gly
                770                 775                 780
Ser Lys Ala Glu Lys Arg Gly Leu Ala Ala Asn Pro Gly Ser Arg
                785                 790                 795
Tyr Asp Asp Leu Val Ser Arg Leu Leu Glu Gln Gly Gly Trp Pro
                800                 805                 810
Gly Glu Leu Leu Ala Ser Trp Glu Ala Gln Asp Ser Ala Glu Arg
                815                 820                 825
Asn Thr Thr Ser Glu Glu Asp Pro Gly Ala Glu Gln Val Leu Leu
                830                 835                 840
His Leu Pro Phe Thr Met Val Thr Glu Gln Arg Arg Leu Asp Asp
                845                 850                 855
Ile Leu Gly Asn Leu Ser Gln Gln Pro Glu Glu Leu Arg Asp Leu
                860                 865                 870
Tyr Ser Lys His Leu Val Ala Gln Leu Ala Gln Glu Ile Phe Arg
                875                 880                 885
Ser His Leu Glu His Gln Asp Thr Leu Leu Lys Pro Ser Glu Arg
                890                 895                 900
Arg Thr Ser Pro Val Thr Leu Ser Pro His Lys His Val Ser Gly
                905                 910                 915
Phe Ser Ser Ser Leu Arg Thr Ser Ser Thr Gly Asp Ala Gly Gly
                920                 925                 930
Gly Ser Arg Arg Pro His Arg Lys Pro Thr Ile Leu Arg Lys Ile
                935                 940                 945
Ser Ala Ala Gln Gln Leu Ser Ala Ser Glu Val Val Thr His Leu
                950                 955                 960
Gly Gln Thr Val Ala Leu Ala Ser Gly Thr Leu Ser Val Leu Leu
                965                 970                 975
His Cys Glu Ala Ile Gly His Pro Arg Pro Thr Ile Ser Trp Ala
                980                 985                 990
Arg Asn Gly Glu Glu Val Gln Phe Ser Asp Arg Ile Leu Leu Gln
                995                1000                1005
Pro Asp Asp Ser Leu Gln Ile Leu Ala Pro Val Glu Ala Asp Val
               1010                1015                1020
Gly Phe Tyr Thr Cys Asn Ala Thr Asn Ala Leu Gly Tyr Asp Ser
               1025                1030                1035
Val Ser Ile Ala Val Thr Leu Ala Gly Lys Pro Leu Val Lys Thr
               1040                1045                1050
Ser Arg Met Thr Val Ile Asn Thr Glu Lys Pro Ala Val Thr Val
               1055                1060                1065
Asp Ile Gly Ser Thr Ile Lys Thr Val Gln Gly Val Asn Val Thr
               1070                1075                1080
Ile Asn Cys Gln Val Ala Gly Val Pro Glu Ala Glu Val Thr Trp
               1085                1090                1095
Phe Arg Asn Lys Ser Lys Leu Gly Ser Pro His His Leu His Glu
               1100                1105                1110
Gly Ser Leu Leu Leu Thr Asn Val Ser Ser Ser Asp Gln Gly Leu
```

1115 1120 1125
Tyr Ser Cys Arg Ala Ala Asn Leu His Gly Glu Leu Thr Glu Ser
1130 1135 1140
Thr Gln Leu Leu Ile Leu Asp Pro Pro Gln Val Pro Thr Gln Leu
1145 1150 1155
Glu Asp Ile Arg Ala Leu Leu Ala Ala Thr Gly Pro Asn Leu Pro
1160 1165 1170
Ser Val Leu Thr Ser Pro Leu Gly Thr Gln Leu Val Leu Gly Pro
1175 1180 1185
Gly Asn Ser Ala Leu Leu Gly Cys Pro Ile Lys Gly His Pro Val
1190 1195 1200
Pro Asn Ile Thr Trp Phe His Gly Gly Gln Pro Ile Val Thr Ala
1205 1210 1215
Thr Gly Leu Thr His His Ile Leu Ala Ala Gly Gln Ile Leu Gln
1220 1225 1230
Val Ala Asn Leu Ser Gly Gly Ser Gln Gly Glu Phe Ser Cys Leu
1235 1240 1245
Ala Gln Asn Glu Ala Gly Val Leu Met Gln Lys Ala Ser Leu Val
1250 1255 1260
Ile Gln Asp Tyr Trp Trp Ser Val Asp Arg Leu Ala Thr Cys Ser
1265 1270 1275
Ala Ser Cys Gly Asn Arg Gly Val Gln Gln Pro Arg Leu Arg Cys
1280 1285 1290
Leu Leu Asn Ser Thr Glu Val Asn Pro Ala His Cys Ala Gly Lys
1295 1300 1305
Val Arg Pro Ala Val Gln Pro Ile Ala Cys Asn Arg Arg Asp Cys
1310 1315 1320
Pro Ser Arg Trp Met Val Thr Ser Trp Ser Ala Cys Thr Arg Ser
1325 1330 1335
Cys Gly Gly Gly Val Gln Thr Arg Arg Val Thr Cys Gln Lys Leu
1340 1345 1350
Lys Ala Ser Gly Ile Ser Thr Pro Val Ser Asn Asp Met Cys Thr
1355 1360 1365
Gln Val Ala Lys Arg Pro Val Asp Thr Gln Ala Cys Asn Gln Gln
1370 1375 1380
Leu Cys Val Glu Trp Ala Phe Ser Ser Trp Gly Gln Cys Asn Gly
1385 1390 1395
Pro Cys Ile Gly Pro His Leu Ala Val Gln His Arg Gln Val Phe
1400 1405 1410
Cys Gln Thr Arg Asp Gly Ile Thr Leu Pro Ser Glu Gln Cys Ser
1415 1420 1425
Ala Leu Pro Arg Pro Val Ser Thr Gln Asn Cys Trp Ser Glu Ala
1430 1435 1440
Cys Ser Val His Trp Arg Val Ser Leu Trp Thr Leu Cys Thr Ala
1445 1450 1455
Thr Cys Gly Asn Tyr Gly Phe Gln Ser Arg Arg Val Glu Cys Val
1460 1465 1470
His Ala Arg Thr Asn Lys Ala Val Pro Glu His Leu Cys Ser Trp
1475 1480 1485
Gly Pro Arg Pro Ala Asn Trp Gln Arg Cys Asn Ile Thr Pro Cys
1490 1495 1500
Glu Asn Met Glu Cys Arg Asp Thr Thr Arg Tyr Cys Glu Lys Val
1505 1510 1515
Lys Gln Leu Lys Leu Cys Gln Leu Ser Gln Phe Lys Ser Arg Cys
1520 1525 1530
Cys Gly Thr Cys Gly Lys Ala
1535

<210> 6
<211> 1120
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472654CD1

<400> 6

```
Met Glu Ile Leu Trp Lys Thr Leu Thr Trp Ile Leu Ser Leu Ile
  1               5                  10                  15
Met Ala Ser Ser Glu Phe His Ser Asp His Arg Leu Ser Tyr Ser
             20                  25                  30
Ser Gln Glu Glu Phe Leu Thr Tyr Leu Glu His Tyr Gln Leu Thr
             35                  40                  45
Ile Pro Ile Arg Val Asp Gln Asn Gly Ala Phe Leu Ser Phe Thr
             50                  55                  60
Val Lys Asn Asp Lys His Ser Arg Arg Arg Arg Ser Met Asp Pro
             65                  70                  75
Ile Asp Pro Gln Gln Ala Val Ser Lys Leu Phe Phe Lys Leu Ser
             80                  85                  90
Ala Tyr Gly Lys His Phe His Leu Asn Leu Thr Leu Asn Thr Asp
             95                  100                 105
Phe Val Ser Lys His Phe Thr Val Glu Tyr Trp Gly Lys Asp Gly
            110                 115                 120
Pro Gln Trp Lys His Asp Phe Leu Asp Asn Cys His Tyr Thr Gly
            125                 130                 135
Tyr Leu Gln Asp Gln Arg Ser Thr Thr Lys Val Ala Leu Ser Asn
            140                 145                 150
Cys Val Gly Leu His Gly Val Ile Ala Thr Glu Asp Glu Glu Tyr
            155                 160                 165
Phe Ile Glu Pro Leu Lys Asn Thr Thr Glu Asp Ser Lys His Phe
            170                 175                 180
Ser Tyr Glu Asn Gly His Pro His Val Ile Tyr Lys Lys Ser Ala
            185                 190                 195
Leu Gln Gln Arg His Leu Tyr Asp His Ser His Cys Gly Val Ser
            200                 205                 210
Asp Phe Thr Arg Ser Gly Lys Pro Trp Trp Leu Asn Asp Thr Ser
            215                 220                 225
Thr Val Ser Tyr Ser Leu Pro Ile Asn Asn Thr His Ile His His
            230                 235                 240
Arg Gln Lys Arg Ser Val Ser Ile Glu Arg Phe Val Glu Thr Leu
            245                 250                 255
Val Val Ala Asp Lys Met Met Val Gly Tyr His Gly Arg Lys Asp
            260                 265                 270
Ile Glu His Tyr Ile Leu Ser Val Met Asn Ile Val Ala Lys Leu
            275                 280                 285
Tyr Arg Asp Ser Ser Leu Gly Asn Val Val Asn Ile Ile Val Ala
            290                 295                 300
Arg Leu Ile Val Leu Thr Glu Asp Gln Pro Asn Leu Glu Ile Asn
            305                 310                 315
His His Ala Asp Lys Ser Leu Asp Ser Phe Cys Lys Trp Gln Lys
            320                 325                 330
Ser Ile Leu Ser His Gln Ser Asp Gly Asn Thr Ile Pro Glu Asn
            335                 340                 345
Gly Ile Ala His His Asp Asn Ala Val Leu Ile Thr Arg Tyr Asp
            350                 355                 360
Ile Cys Thr Tyr Lys Asn Lys Pro Cys Gly Thr Leu Gly Leu Ala
            365                 370                 375
Ser Val Ala Gly Met Cys Glu Pro Glu Arg Ser Cys Ser Ile Asn
            380                 385                 390
Glu Asp Ile Gly Leu Gly Ser Ala Phe Thr Ile Ala His Glu Ile
            395                 400                 405
Gly His Asn Phe Gly Met Asn His Asp Gly Ile Gly Asn Ser Cys
            410                 415                 420
Gly Thr Lys Gly His Glu Ala Ala Lys Leu Met Ala Ala His Ile
            425                 430                 435
```

```
Thr Ala Asn Thr Asn Pro Phe Ser Trp Ser Ala Cys Ser Arg Asp
                440                 445                 450
Tyr Ile Thr Ser Phe Leu Asp Ser Gly Arg Gly Thr Cys Leu Asp
                455                 460                 465
Asn Glu Pro Pro Lys Arg Asp Phe Leu Tyr Pro Ala Val Ala Pro
                470                 475                 480
Gly Gln Val Tyr Asp Ala Asp Glu Gln Cys Arg Phe Gln Tyr Gly
                485                 490                 495
Ala Thr Ser Arg Gln Cys Lys Tyr Gly Glu Val Cys Arg Glu Leu
                500                 505                 510
Trp Cys Leu Ser Lys Ser Asn Arg Cys Val Thr Asn Ser Ile Pro
                515                 520                 525
Ala Ala Glu Gly Thr Leu Cys Gln Thr Gly Asn Ile Glu Lys Gly
                530                 535                 540
Trp Cys Tyr Gln Gly Asp Cys Val Pro Phe Gly Thr Trp Pro Gln
                545                 550                 555
Ser Ile Asp Gly Gly Trp Gly Pro Trp Ser Leu Trp Gly Glu Cys
                560                 565                 570
Ser Arg Thr Cys Gly Gly Gly Val Ser Ser Ser Leu Arg His Cys
                575                 580                 585
Asp Ser Pro Ala Phe Phe Arg Pro Ser Gly Gly Gly Lys Tyr Cys
                590                 595                 600
Leu Gly Glu Arg Lys Arg Tyr Arg Ser Cys Asn Thr Asp Pro Cys
                605                 610                 615
Pro Leu Gly Ser Arg Asp Phe Arg Glu Lys Gln Cys Ala Asp Phe
                620                 625                 630
Asp Asn Met Pro Phe Arg Gly Lys Tyr Tyr Asn Trp Lys Pro Tyr
                635                 640                 645
Thr Gly Gly Gly Val Lys Pro Cys Ala Leu Asn Cys Leu Ala Glu
                650                 655                 660
Gly Tyr Asn Phe Tyr Thr Glu Arg Ala Pro Ala Val Ile Asp Gly
                665                 670                 675
Thr Gln Cys Asn Ala Asp Ser Leu Asp Ile Cys Ile Asn Gly Glu
                680                 685                 690
Cys Lys His Val Gly Cys Asp Asn Ile Leu Gly Ser Asp Ala Arg
                695                 700                 705
Glu Asp Arg Cys Arg Val Cys Gly Gly Asp Gly Ser Thr Cys Asp
                710                 715                 720
Ala Ile Glu Gly Phe Phe Asn Asp Ser Leu Pro Arg Gly Gly Tyr
                725                 730                 735
Met Glu Val Val Gln Ile Pro Arg Gly Ser Val His Ile Glu Val
                740                 745                 750
Arg Glu Val Ala Met Ser Lys Asn Tyr Ile Ala Leu Lys Ser Glu
                755                 760                 765
Gly Asp Asp Tyr Tyr Ile Asn Gly Ala Trp Thr Ile Asp Trp Pro
                770                 775                 780
Arg Lys Phe Asp Val Ala Gly Thr Ala Phe His Tyr Lys Arg Pro
                785                 790                 795
Thr Asp Glu Pro Glu Ser Leu Glu Ala Leu Gly Pro Thr Ser Glu
                800                 805                 810
Asn Leu Ile Val Met Val Leu Leu Gln Glu Gln Asn Leu Gly Ile
                815                 820                 825
Arg Tyr Lys Phe Asn Val Pro Ile Thr Arg Thr Gly Ser Gly Asp
                830                 835                 840
Asn Glu Val Gly Phe Thr Trp Asn His Gln Pro Trp Ser Glu Cys
                845                 850                 855
Ser Ala Thr Cys Ala Gly Gly Val Gln Arg Gln Glu Val Val Cys
                860                 865                 870
Lys Arg Leu Asp Asp Asn Ser Ile Val Gln Asn Asn Tyr Cys Asp
                875                 880                 885
Pro Asp Ser Lys Pro Pro Glu Asn Gln Arg Ala Cys Asn Thr Glu
                890                 895                 900
Pro Cys Pro Pro Glu Trp Phe Ile Gly Asp Trp Leu Glu Cys Ser
```

905 910 915
Lys Thr Cys Asp Gly Gly Met Arg Thr Arg Ala Val Leu Cys Ile
920 925 930
Arg Lys Ile Gly Pro Ser Glu Glu Glu Thr Leu Asp Tyr Ser Gly
935 940 945
Cys Leu Thr His Arg Pro Val Glu Lys Glu Pro Cys Asn Asn Gln
950 955 960
Ser Cys Pro Pro Gln Trp Val Ala Leu Asp Trp Ser Glu Cys Thr
965 970 975
Pro Lys Cys Gly Pro Gly Phe Lys His Arg Ile Val Leu Cys Lys
980 985 990
Ser Ser Asp Leu Ser Lys Thr Phe Pro Ala Ala Gln Cys Pro Glu
995 1000 1005
Glu Ser Lys Pro Pro Val Arg Ile Arg Cys Ser Leu Gly Arg Cys
1010 1015 1020
Pro Pro Pro Arg Trp Val Thr Gly Asp Trp Gly Gln Cys Ser Ala
1025 1030 1035
Gln Cys Gly Leu Gly Gln Gln Met Arg Thr Val Gln Cys Leu Ser
1040 1045 1050
Tyr Thr Gly Gln Ala Ser Ser Asp Cys Leu Glu Thr Val Arg Pro
1055 1060 1065
Pro Ser Met Gln Gln Cys Glu Ser Lys Cys Asp Ser Thr Pro Ile
1070 1075 1080
Ser Asn Thr Glu Glu Cys Lys Asp Val Asn Lys Val Ala Tyr Cys
1085 1090 1095
Pro Leu Val Leu Lys Phe Lys Phe Cys Ser Arg Ala Tyr Phe Arg
1100 1105 1110
Gln Met Cys Cys Lys Thr Cys Gln Gly His
1115 1120

<210> 7
<211> 328
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480224CD1

<400> 7
Met Gly Pro Ala Gly Cys Ala Phe Thr Leu Leu Leu Leu Leu Gly
1 5 10 15
Ile Ser Val Cys Gly Gln Pro Val Tyr Ser Ser Arg Val Val Gly
20 25 30
Gly Gln Asp Ala Ala Ala Gly Arg Trp Pro Trp Gln Val Ser Leu
35 40 45
His Phe Asp His Asn Phe Ile Tyr Gly Gly Ser Leu Val Ser Glu
50 55 60
Arg Leu Ile Leu Thr Ala Ala His Cys Ile Gln Pro Thr Trp Thr
65 70 75
Thr Phe Ser Tyr Thr Val Trp Leu Gly Ser Ile Thr Val Gly Asp
80 85 90
Ser Arg Lys Arg Val Lys Tyr Tyr Val Ser Lys Ile Val Ile His
95 100 105
Pro Lys Tyr Gln Asp Thr Thr Ala Asp Val Ala Leu Leu Lys Leu
110 115 120
Ser Ser Gln Val Thr Phe Thr Ser Ala Ile Leu Pro Ile Cys Leu
125 130 135
Pro Ser Val Thr Lys Gln Leu Ala Ile Pro Pro Phe Cys Trp Val
140 145 150
Thr Gly Trp Gly Lys Val Lys Glu Ser Ser Asp Arg Asp Tyr His
155 160 165
Ser Ala Leu Gln Glu Ala Glu Val Pro Ile Ile Asp Arg Gln Ala

```
                170                 175                 180
Cys Glu Gln Leu Tyr Asn Pro Ile Gly Ile Phe Leu Pro Ala Leu
                185                 190                 195
Glu Pro Val Ile Lys Glu Asp Lys Ile Cys Ala Gly Asp Thr Gln
                200                 205                 210
Asn Met Lys Asp Ser Cys Lys Gly Asp Ser Gly Gly Pro Leu Ser
                215                 220                 225
Cys His Ile Asp Gly Val Trp Ile Gln Thr Gly Val Val Ser Trp
                230                 235                 240
Gly Leu Glu Cys Gly Lys Ser Leu Pro Gly Val Tyr Thr Asn Val
                245                 250                 255
Ile Tyr Tyr Gln Lys Trp Ile Asn Ala Thr Ile Ser Arg Ala Asn
                260                 265                 270
Asn Leu Asp Phe Ser Asp Phe Leu Phe Pro Ile Val Leu Leu Ser
                275                 280                 285
Leu Ala Leu Leu Arg Pro Ser Cys Ala Phe Gly Pro Asn Thr Ile
                290                 295                 300
His Arg Val Gly Thr Val Ala Glu Ala Val Ala Cys Ile Gln Gly
                305                 310                 315
Trp Glu Glu Asn Ala Trp Arg Phe Ser Pro Arg Gly Arg
                320                 325
```

<210> 8
<211> 425
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481056CD1

<400> 8

```
Met Met Tyr Ala Pro Val Glu Phe Ser Glu Ala Glu Phe Ser Arg
  1                 5                  10                  15
Ala Glu Tyr Gln Arg Lys Gln Gln Phe Trp Asp Ser Val Arg Leu
                 20                  25                  30
Ala Leu Phe Thr Leu Ala Ile Val Ala Ile Ile Gly Ile Ala Ile
                 35                  40                  45
Gly Ile Val Thr His Phe Val Val Glu Asp Asp Lys Ser Phe Tyr
                 50                  55                  60
Tyr Leu Ala Ser Phe Lys Val Thr Asn Ile Lys Tyr Lys Glu Asn
                 65                  70                  75
Tyr Gly Ile Arg Ser Ser Arg Glu Phe Ile Glu Arg Ser His Gln
                 80                  85                  90
Ile Glu Arg Met Met Ser Arg Ile Phe Arg His Ser Ser Val Gly
                 95                 100                 105
Gly Arg Phe Ile Lys Ser His Val Ile Lys Leu Ser Pro Asp Glu
                110                 115                 120
Gln Gly Val Asp Ile Leu Ile Val Leu Ile Phe Arg Tyr Pro Ser
                125                 130                 135
Thr Asp Ser Ala Glu Gln Ile Lys Lys Lys Ile Glu Lys Ala Leu
                140                 145                 150
Tyr Gln Ser Leu Lys Thr Lys Gln Leu Ser Leu Thr Ile Asn Lys
                155                 160                 165
Pro Ser Phe Arg Leu Thr Arg Cys Gly Ile Arg Met Thr Ser Ser
                170                 175                 180
Asn Met Pro Leu Pro Ala Ser Ser Ser Thr Gln Arg Ile Val Gln
                185                 190                 195
Gly Arg Glu Thr Ala Met Glu Gly Glu Trp Pro Trp Gln Ala Ser
                200                 205                 210
Leu Gln Leu Ile Gly Ser Gly His Gln Cys Gly Ala Ser Leu Ile
                215                 220                 225
Ser Asn Thr Trp Leu Leu Thr Ala Ala His Cys Phe Trp Lys Asn
```

230 235 240
Lys Asp Pro Thr Gln Trp Ile Ala Thr Phe Gly Ala Thr Ile Thr
245 250 255
Pro Pro Ala Val Lys Arg Asn Val Arg Lys Ile Ile Leu His Glu
260 265 270
Asn Tyr His Arg Glu Thr Asn Glu Asn Asp Ile Ala Leu Val Gln
275 280 285
Leu Ser Thr Gly Val Glu Phe Ser Asn Ile Val Gln Arg Val Cys
290 295 300
Leu Pro Asp Ser Ser Ile Lys Leu Pro Pro Lys Thr Ser Val Phe
305 310 315
Val Thr Gly Phe Gly Ser Ile Val Asp Asp Gly Pro Ile Gln Asn
320 325 330
Thr Leu Arg Gln Ala Arg Val Glu Thr Ile Ser Thr Asp Val Cys
335 340 345
Asn Arg Lys Asp Val Tyr Asp Gly Leu Ile Thr Pro Gly Met Leu
350 355 360
Cys Ala Gly Phe Met Glu Gly Lys Ile Asp Ala Cys Lys Gly Asp
365 370 375
Ser Gly Gly Pro Leu Val Tyr Asp Asn His Asp Ile Trp Tyr Ile
380 385 390
Val Gly Ile Val Ser Trp Gly Gln Ser Cys Ala Leu Pro Lys Lys
395 400 405
Pro Gly Val Tyr Thr Arg Val Thr Lys Tyr Arg Asp Trp Ile Ala
410 415 420
Ser Lys Thr Gly Met
425

<210> 9
<211> 1103
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3750264CD1

<400> 9
Met Ala Pro Ala Cys Gln Ile Leu Arg Trp Ala Leu Ala Leu Gly
1 5 10 15
Leu Gly Leu Met Phe Glu Val Thr His Ala Phe Arg Ser Gln Asp
20 25 30
Glu Phe Leu Ser Ser Leu Glu Ser Tyr Glu Ile Ala Phe Pro Thr
35 40 45
Arg Val Asp His Asn Gly Ala Leu Leu Ala Phe Ser Pro Pro Pro
50 55 60
Pro Arg Arg Gln Arg Arg Gly Thr Gly Ala Thr Ala Glu Ser Arg
65 70 75
Leu Phe Tyr Lys Val Ala Ser Pro Ser Thr His Phe Leu Leu Asn
80 85 90
Leu Thr Arg Ser Ser Arg Leu Leu Ala Gly His Val Ser Val Glu
95 100 105
Tyr Trp Thr Arg Glu Gly Leu Ala Trp Gln Arg Ala Ala Arg Pro
110 115 120
His Cys Leu Tyr Ala Gly His Leu Gln Gly Gln Ala Ser Ser Ser
125 130 135
His Val Ala Ile Ser Thr Cys Gly Gly Leu His Gly Leu Ile Val
140 145 150
Ala Asp Glu Glu Glu Tyr Leu Ile Glu Pro Leu His Gly Gly Pro
155 160 165
Lys Gly Ser Arg Ser Pro Glu Glu Ser Gly Pro His Val Val Tyr
170 175 180
Lys Arg Ser Ser Leu Arg His Pro His Leu Asp Thr Ala Cys Gly

```
                185                 190                 195
Val Arg Asp Glu Lys Pro Trp Lys Gly Arg Pro Trp Trp Leu Arg
            200                 205                 210
Thr Leu Lys Pro Pro Pro Ala Arg Pro Leu Gly Asn Glu Thr Glu
        215                 220                 225
Arg Gly Gln Pro Gly Leu Lys Arg Ser Val Ser Arg Glu Arg Tyr
    230                 235                 240
Val Glu Thr Leu Val Val Ala Asp Lys Met Met Val Ala Tyr His
245                 250                 255
Gly Arg Arg Asp Val Glu Gln Tyr Val Leu Ala Val Met Asn Ile
            260                 265                 270
Val Ala Lys Leu Phe Gln Asp Ser Ser Leu Gly Ser Thr Val Asn
        275                 280                 285
Ile Leu Val Thr Arg Leu Ile Leu Leu Thr Glu Asp Gln Pro Thr
    290                 295                 300
Leu Glu Ile Thr His His Ala Gly Lys Ser Leu Asp Ser Phe Cys
305                 310                 315
Lys Trp Gln Lys Ser Ile Val Asn His Ser Gly His Gly Asn Ala
            320                 325                 330
Ile Pro Glu Asn Gly Val Ala Asn His Asp Thr Ala Val Leu Ile
        335                 340                 345
Thr Arg Tyr Asp Ile Cys Ile Tyr Lys Asn Lys Pro Cys Gly Thr
    350                 355                 360
Leu Gly Leu Ala Pro Val Gly Gly Met Cys Glu Arg Glu Arg Ser
365                 370                 375
Cys Ser Val Asn Glu Asp Ile Gly Leu Ala Thr Ala Phe Thr Ile
            380                 385                 390
Ala His Glu Ile Gly His Thr Phe Gly Met Asn His Asp Gly Val
        395                 400                 405
Gly Asn Ser Cys Gly Ala Arg Gly Gln Asp Pro Ala Lys Leu Met
    410                 415                 420
Ala Ala His Ile Thr Met Lys Thr Asn Pro Phe Val Trp Ser Ser
425                 430                 435
Cys Ser Arg Asp Tyr Ile Thr Ser Phe Leu Asp Ser Gly Leu Gly
            440                 445                 450
Leu Cys Leu Asn Asn Arg Pro Pro Arg Gln Asp Phe Val Tyr Pro
        455                 460                 465
Thr Val Ala Pro Gly Gln Ala Tyr Asp Ala Asp Glu Gln Cys Arg
    470                 475                 480
Phe Gln His Gly Val Lys Ser Arg Gln Cys Lys Tyr Gly Glu Val
485                 490                 495
Cys Ser Glu Leu Trp Cys Leu Ser Lys Ser Asn Arg Cys Ile Thr
            500                 505                 510
Asn Ser Ile Pro Ala Ala Glu Gly Thr Leu Cys Gln Thr His Thr
        515                 520                 525
Ile Asp Lys Gly Trp Cys Tyr Lys Arg Val Cys Val Pro Phe Gly
    530                 535                 540
Ser Arg Pro Glu Gly Val Asp Gly Ala Trp Gly Pro Trp Thr Pro
545                 550                 555
Trp Gly Asp Cys Ser Arg Thr Cys Gly Gly Gly Val Ser Ser Ser
            560                 565                 570
Ser Arg His Cys Asp Ser Pro Arg Pro Thr Ile Gly Gly Lys Tyr
        575                 580                 585
Cys Leu Gly Glu Arg Arg Arg His Arg Ser Cys Asn Thr Asp Asp
    590                 595                 600
Cys Pro Pro Gly Ser Gln Asp Phe Arg Glu Val Gln Cys Ser Glu
605                 610                 615
Phe Asp Ser Ile Pro Phe Arg Gly Lys Phe Tyr Lys Trp Lys Thr
            620                 625                 630
Tyr Arg Gly Gly Gly Val Lys Ala Cys Ser Leu Thr Cys Leu Ala
        635                 640                 645
Glu Gly Phe Asn Phe Tyr Thr Glu Arg Ala Ala Ala Val Val Asp
    650                 655                 660
```

Gly Thr Pro Cys Arg Pro Asp Thr Val Asp Ile Cys Val Ser Gly
665 670 675
Glu Cys Lys His Val Gly Cys Asp Arg Val Leu Gly Ser Asp Leu
680 685 690
Arg Glu Asp Lys Cys Arg Val Cys Gly Gly Asp Gly Ser Ala Cys
695 700 705
Glu Thr Ile Glu Gly Val Phe Ser Pro Ala Ser Pro Gly Ala Gly
710 715 720
Tyr Glu Asp Val Val Trp Ile Pro Lys Gly Ser Val His Ile Phe
725 730 735
Ile Gln Asp Leu Asn Leu Ser Leu Ser His Leu Ala Leu Lys Gly
740 745 750
Asp Gln Glu Ser Leu Leu Leu Glu Gly Leu Pro Gly Thr Pro Gln
755 760 765
Pro His Arg Leu Pro Leu Ala Gly Thr Thr Phe Gln Leu Arg Gln
770 775 780
Gly Pro Asp Gln Val Gln Ser Leu Glu Ala Leu Gly Pro Ile Asn
785 790 795
Ala Ser Leu Ile Val Met Val Leu Ala Arg Thr Glu Leu Pro Ala
800 805 810
Leu Arg Tyr Arg Phe Asn Ala Pro Ile Ala Arg Asp Ser Leu Pro
815 820 825
Pro Tyr Ser Trp His Tyr Ala Pro Trp Thr Lys Cys Ser Ala Gln
830 835 840
Cys Ala Gly Gly Ser Gln Val Gln Ala Val Glu Cys Arg Asn Gln
845 850 855
Leu Asp Ser Ser Ala Val Ala Pro His Tyr Cys Ser Ala His Ser
860 865 870
Lys Leu Pro Lys Arg Gln Arg Ala Cys Asn Thr Glu Pro Cys Pro
875 880 885
Pro Asp Trp Val Val Gly Asn Trp Ser Leu Cys Ser Arg Ser Cys
890 895 900
Asp Ala Gly Val Arg Ser Arg Ser Val Val Cys Gln Arg Arg Val
905 910 915
Ser Ala Ala Glu Glu Lys Ala Leu Asp Asp Ser Ala Cys Pro Gln
920 925 930
Pro Arg Pro Pro Val Leu Glu Ala Cys His Gly Pro Thr Cys Pro
935 940 945
Pro Glu Trp Ala Ala Leu Asp Trp Ser Glu Cys Thr Pro Ser Cys
950 955 960
Gly Pro Gly Leu Arg His Arg Val Val Leu Cys Lys Ser Ala Asp
965 970 975
His Arg Ala Thr Leu Pro Pro Ala His Cys Ser Pro Ala Ala Lys
980 985 990
Pro Pro Ala Thr Met Arg Cys Asn Leu Arg Arg Cys Pro Pro Ala
995 1000 1005
Arg Trp Val Ala Gly Glu Trp Gly Glu Cys Ser Ala Gln Cys Gly
1010 1015 1020
Val Gly Gln Arg Gln Arg Ser Val Arg Cys Thr Ser His Thr Gly
1025 1030 1035
Gln Ala Ser His Glu Cys Thr Glu Ala Leu Arg Pro Pro Thr Thr
1040 1045 1050
Gln Gln Cys Glu Ala Lys Cys Asp Ser Pro Thr Pro Gly Asp Gly
1055 1060 1065
Pro Glu Glu Cys Lys Asp Val Asn Lys Val Ala Tyr Cys Pro Leu
1070 1075 1080
Val Leu Lys Phe Gln Phe Cys Ser Arg Ala Tyr Phe Arg Gln Met
1085 1090 1095
Cys Cys Lys Thr Cys Gln Gly His
1100

<210> 10
<211> 83

<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1749735CD1

<400> 10
Met Phe Leu Thr Phe Val Val Leu Thr Ser Leu Thr Pro Leu Trp
1 5 10 15
Ser Gly Asn Ala Cys Val Arg Ser Ile Asp Ala Phe Pro Pro Gln
20 25 30
Gln Phe His His Ala Ile Phe Thr Leu Gly Tyr Asp Ser Pro Ala
35 40 45
Lys Ser Ser Val His Gln Met Tyr Thr Ser Ile Val Gly Pro Arg
50 55 60
Cys Leu Ser Ala Thr His Cys Phe Ser Val Phe Leu Leu Leu Lys
65 70 75
Cys Ser Glu Met Asn Pro Ser Asn
80

<210> 11
<211> 1274
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473634CD1

<400> 11
Met Val Thr Ile Cys Leu Val Thr Ala Trp Thr Gly Leu Ser Trp
1 5 10 15
Ser Tyr His Leu Arg Ser His Ile Leu Glu Thr Pro Leu Ile Val
20 25 30
Glu Asn Arg Asn Ile Trp Thr Ser Asn Glu Arg Asp Arg Gly Ser
35 40 45
Gln Ser Val Gly Thr Thr Gly Ile Ser His Arg Ala Lys Pro Val
50 55 60
Ser Cys Phe Leu Lys Tyr Lys Ala Thr Glu Gly Ala Cys Gly Gly
65 70 75
Thr Leu Arg Gly Thr Ser Ser Ser Ile Ser Ser Pro His Phe Pro
80 85 90
Ser Glu Tyr Glu Asn Asn Ala Asp Cys Thr Trp Thr Ile Leu Ala
95 100 105
Glu Pro Gly Asp Thr Ile Ala Leu Val Phe Thr Asp Phe Gln Leu
110 115 120
Glu Glu Gly Tyr Asp Phe Leu Glu Ile Ser Gly Thr Glu Ala Pro
125 130 135
Ser Ile Trp Leu Thr Gly Met Asn Leu Pro Ser Pro Val Ile Ser
140 145 150
Ser Lys Asn Trp Leu Arg Leu His Phe Thr Ser Asp Ser Asn His
155 160 165
Arg Arg Lys Gly Phe Asn Ala Gln Phe Gln Val Lys Lys Ala Ile
170 175 180
Glu Leu Lys Ser Arg Gly Val Lys Met Leu Pro Ser Lys Asp Gly
185 190 195
Ser His Lys Asn Ser Val Leu Ser Gln Gly Gly Val Ala Leu Val
200 205 210
Ser Asp Met Cys Pro Asp Pro Gly Ile Pro Glu Asn Gly Arg Arg
215 220 225
Ala Gly Ser Asp Phe Arg Val Gly Ala Asn Val Gln Phe Ser Cys
230 235 240

```
Glu Asp Asn Tyr Val Leu Gln Gly Ser Lys Ser Ile Thr Cys Gln
                245                 250                 255
Arg Val Thr Glu Thr Leu Ala Ala Trp Ser Asp His Arg Pro Ile
            260                 265                 270
Cys Arg Ala Arg Thr Cys Gly Ser Asn Leu Arg Gly Pro Ser Gly
        275                 280                 285
Val Ile Thr Ser Pro Asn Tyr Pro Val Gln Tyr Glu Asp Asn Ala
    290                 295                 300
His Cys Val Trp Val Ile Thr Thr Thr Asp Pro Asp Lys Val Ile
305                 310                 315
Lys Leu Ala Phe Glu Glu Phe Glu Leu Glu Arg Gly Tyr Asp Thr
            320                 325                 330
Leu Thr Val Gly Asp Ala Gly Lys Val Gly Asp Thr Arg Ser Val
        335                 340                 345
Leu Tyr Val Leu Thr Gly Ser Ser Val Pro Asp Leu Ile Val Ser
    350                 355                 360
Met Ser Asn Gln Met Trp Leu His Leu Gln Ser Asp Asp Ser Ile
365                 370                 375
Gly Ser Pro Gly Phe Lys Ala Val Tyr Gln Glu Ile Glu Lys Gly
            380                 385                 390
Gly Cys Gly Asp Pro Gly Ile Pro Ala Tyr Gly Lys Arg Thr Gly
        395                 400                 405
Ser Ser Phe Leu His Gly Asp Thr Leu Thr Phe Glu Cys Pro Ala
    410                 415                 420
Ala Phe Glu Leu Val Gly Glu Arg Val Ile Thr Cys Gln Gln Asn
425                 430                 435
Asn Gln Trp Ser Gly Asn Lys Pro Ser Cys Val Phe Ser Cys Phe
            440                 445                 450
Phe Asn Phe Thr Ala Ser Ser Gly Ile Ile Leu Ser Pro Asn Tyr
        455                 460                 465
Pro Glu Glu Tyr Gly Asn Asn Met Asn Cys Val Trp Leu Ile Ile
    470                 475                 480
Ser Glu Pro Gly Ser Arg Ile His Leu Ile Phe Asn Asp Phe Asp
485                 490                 495
Val Glu Pro Gln Phe Asp Phe Leu Ala Val Lys Asp Asp Gly Ile
            500                 505                 510
Ser Asp Ile Thr Val Leu Gly Thr Phe Ser Gly Asn Glu Val Pro
        515                 520                 525
Ser Gln Leu Ala Ser Ser Gly His Ile Val Arg Leu Glu Phe Gln
    530                 535                 540
Ser Asp His Ser Thr Thr Gly Arg Gly Phe Asn Ile Thr Tyr Thr
545                 550                 555
Thr Phe Gly Gln Asn Glu Cys His Asp Pro Gly Ile Pro Ile Asn
            560                 565                 570
Gly Arg Arg Phe Gly Asp Arg Phe Leu Leu Gly Ser Ser Val Ser
        575                 580                 585
Phe His Cys Asp Asp Gly Phe Val Lys Thr Gln Gly Ser Glu Ser
    590                 595                 600
Ile Thr Cys Ile Leu Gln Asp Gly Asn Val Val Trp Ser Ser Thr
605                 610                 615
Val Pro Arg Cys Glu Ala Pro Cys Gly Gly His Leu Thr Ala Ser
            620                 625                 630
Ser Gly Val Ile Leu Pro Pro Gly Trp Pro Gly Tyr Tyr Lys Asp
        635                 640                 645
Ser Leu His Cys Glu Trp Ile Ile Glu Ala Lys Pro Gly His Ser
    650                 655                 660
Ile Lys Ile Thr Phe Asp Arg Phe Gln Thr Glu Val Asn Tyr Asp
665                 670                 675
Thr Leu Glu Val Arg Asp Gly Pro Ala Ser Ser Ser Pro Leu Ile
            680                 685                 690
Gly Glu Tyr His Gly Thr Gln Ala Pro Gln Phe Leu Ile Ser Thr
        695                 700                 705
Gly Asn Phe Met Tyr Leu Leu Phe Thr Thr Asp Asn Ser Arg Ser
```

710 715 720
Ser Ile Gly Phe Leu Ile His Tyr Glu Ser Val Thr Leu Glu Ser
725 730 735
Asp Ser Cys Leu Asp Pro Gly Ile Pro Val Asn Gly His Arg His
740 745 750
Gly Gly Asp Phe Gly Ile Arg Ser Thr Val Thr Phe Ser Cys Asp
755 760 765
Pro Gly Tyr Thr Leu Ser Asp Asp Glu Pro Leu Val Cys Glu Arg
770 775 780
Asn His Gln Trp Asn His Ala Leu Pro Ser Cys Asp Ala Leu Cys
785 790 795
Gly Gly Tyr Ile Gln Gly Lys Ser Gly Thr Val Leu Ser Pro Gly
800 805 810
Phe Pro Asp Phe Tyr Pro Asn Ser Leu Asn Cys Thr Trp Thr Ile
815 820 825
Glu Val Ser His Gly Lys Gly Val Gln Met Ile Phe His Thr Phe
830 835 840
His Leu Glu Ser Ser His Asp Tyr Leu Leu Ile Thr Glu Asp Gly
845 850 855
Ser Phe Ser Glu Pro Val Ala Arg Leu Thr Gly Ser Val Leu Pro
860 865 870
His Thr Ile Lys Ala Gly Leu Phe Gly Asn Phe Thr Ala Gln Leu
875 880 885
Arg Phe Ile Ser Asp Phe Ser Ile Ser Tyr Glu Gly Phe Asn Ile
890 895 900
Thr Phe Ser Glu Tyr Asp Leu Glu Pro Cys Asp Asp Pro Gly Val
905 910 915
Pro Ala Phe Ser Arg Arg Ile Gly Phe His Phe Gly Val Gly Asp
920 925 930
Ser Leu Thr Phe Ser Cys Phe Leu Gly Tyr Arg Leu Glu Gly Ala
935 940 945
Thr Lys Leu Thr Cys Leu Gly Gly Gly Arg Arg Val Trp Ser Ala
950 955 960
Pro Leu Pro Arg Cys Val Ala Glu Cys Gly Ala Ser Val Lys Gly
965 970 975
Asn Glu Gly Thr Leu Leu Ser Pro Asn Phe Pro Ser Asn Tyr Asp
980 985 990
Asn Asn His Glu Cys Ile Tyr Lys Ile Glu Thr Glu Ala Gly Lys
995 1000 1005
Gly Ile His Leu Arg Thr Arg Ser Phe Gln Leu Phe Glu Gly Asp
1010 1015 1020
Thr Leu Lys Val Tyr Asp Gly Lys Asp Ser Ser Ser Arg Pro Leu
1025 1030 1035
Gly Thr Phe Thr Lys Asn Glu Leu Leu Gly Leu Ile Leu Asn Ser
1040 1045 1050
Thr Ser Asn His Leu Trp Leu Glu Phe Asn Thr Asn Gly Ser Asp
1055 1060 1065
Thr Asp Gln Gly Phe Gln Leu Thr Tyr Thr Ser Phe Asp Leu Val
1070 1075 1080
Lys Cys Glu Asp Pro Gly Ile Pro Asn Tyr Gly Tyr Arg Ile Arg
1085 1090 1095
Asp Glu Gly His Phe Thr Asp Thr Val Val Leu Tyr Ser Cys Asn
1100 1105 1110
Pro Gly Tyr Ala Met His Gly Ser Asn Thr Leu Thr Cys Leu Ser
1115 1120 1125
Gly Asp Arg Arg Val Trp Asp Lys Pro Leu Pro Ser Cys Ile Ala
1130 1135 1140
Glu Cys Gly Gly Gln Ile His Ala Ala Thr Ser Gly Arg Ile Leu
1145 1150 1155
Ser Pro Gly Tyr Pro Ala Pro Tyr Asp Asn Asn Leu His Cys Thr
1160 1165 1170
Trp Ile Ile Glu Ala Asp Pro Gly Lys Thr Ile Ser Leu His Phe
1175 1180 1185

```
Ile Val Phe Asp Thr Glu Met Ala His Asp Ile Leu Lys Val Trp
                1190                1195                1200
Asp Gly Pro Val Asp Ser Asp Ile Leu Leu Lys Glu Trp Ser Gly
                1205                1210                1215
Ser Ala Leu Pro Glu Asp Ile His Ser Thr Phe Asn Ser Leu Thr
                1220                1225                1230
Leu Gln Phe Asp Ser Asp Phe Phe Ile Ser Lys Ser Gly Phe Ser
                1235                1240                1245
Ile Gln Phe Ser Arg Ser Gln Ala Gly Thr Arg Arg Arg Trp Ser
                1250                1255                1260
Asp His Pro Lys Ala Ser His Ser Ala Thr Leu His Lys Met
                1265                1270
```

<210> 12
<211> 243
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4767844CD1

<400> 12

```
Met Gln Phe Arg Leu Phe Ser Phe Ala Leu Ile Ile Leu Asn Cys
  1               5                  10                  15
Met Asp Tyr Ser His Cys Gln Gly Asn Arg Trp Arg Arg Ser Lys
                20                  25                  30
Arg Ala Ser Tyr Val Ser Asn Pro Ile Cys Lys Gly Cys Leu Ser
                35                  40                  45
Cys Ser Lys Asp Asn Gly Cys Ser Arg Cys Gln Gln Lys Leu Phe
                50                  55                  60
Phe Phe Leu Arg Arg Glu Gly Met Arg Gln Tyr Gly Glu Cys Leu
                65                  70                  75
His Ser Cys Pro Ser Gly Tyr Tyr Gly His Arg Ala Pro Asp Met
                80                  85                  90
Asn Arg Cys Ala Arg Cys Arg Ile Glu Asn Cys Asp Ser Cys Phe
                95                  100                 105
Ser Lys Asp Phe Cys Thr Lys Cys Lys Val Gly Phe Tyr Leu His
                110                 115                 120
Arg Gly Arg Cys Phe Asp Glu Cys Pro Asp Gly Phe Ala Pro Leu
                125                 130                 135
Glu Glu Thr Met Glu Cys Val Glu Gly Cys Glu Val Gly His Trp
                140                 145                 150
Ser Glu Trp Gly Thr Cys Ser Arg Asn Asn Arg Thr Cys Gly Phe
                155                 160                 165
Lys Trp Gly Leu Glu Thr Arg Thr Arg Gln Ile Val Lys Lys Pro
                170                 175                 180
Val Lys Asp Thr Ile Pro Cys Pro Thr Ile Ala Glu Ser Arg Arg
                185                 190                 195
Cys Lys Met Thr Met Arg His Cys Pro Gly Gly Lys Arg Thr Pro
                200                 205                 210
Lys Ala Lys Glu Lys Arg Asn Lys Lys Lys Lys Arg Lys Leu Ile
                215                 220                 225
Glu Arg Ala Gln Glu Gln His Ser Val Phe Leu Ala Thr Asp Arg
                230                 235                 240
Ala Asn Gln
```

<210> 13
<211> 672
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7487584CD1

<400> 13

```
Met Glu Cys Cys Arg Arg Ala Thr Pro Gly Thr Leu Leu Leu Phe
  1               5                  10                  15
Leu Ala Phe Leu Leu Leu Ser Ser Arg Thr Ala Arg Ser Glu Glu
             20                  25                  30
Asp Arg Asp Gly Leu Trp Asp Ala Trp Gly Pro Trp Ser Glu Cys
             35                  40                  45
Ser Arg Thr Cys Gly Gly Gly Ala Ser Tyr Ser Leu Arg Arg Cys
             50                  55                  60
Leu Ser Ser Lys Ser Cys Glu Gly Arg Asn Ile Arg Tyr Arg Thr
             65                  70                  75
Cys Ser Asn Val Asp Cys Pro Pro Glu Ala Gly Asp Phe Arg Ala
             80                  85                  90
Gln Gln Cys Ser Ala His Asn Asp Val Lys His His Gly Gln Phe
             95                 100                 105
Tyr Glu Trp Leu Pro Val Ser Asn Asp Pro Asp Asn Pro Cys Ser
            110                 115                 120
Leu Lys Cys Gln Ala Lys Gly Thr Thr Leu Val Val Glu Leu Ala
            125                 130                 135
Pro Lys Val Leu Asp Gly Thr Arg Cys Tyr Thr Glu Ser Leu Asp
            140                 145                 150
Met Cys Ile Ser Gly Leu Cys Gln Ile Val Gly Cys Asp His Gln
            155                 160                 165
Leu Gly Ser Thr Val Lys Glu Asp Asn Cys Gly Val Cys Asn Gly
            170                 175                 180
Asp Gly Ser Thr Cys Arg Leu Val Arg Gly Gln Tyr Lys Ser Gln
            185                 190                 195
Leu Ser Ala Thr Lys Ser Asp Asp Thr Val Val Ala Ile Pro Tyr
            200                 205                 210
Gly Ser Arg His Ile Arg Leu Val Leu Lys Gly Pro Asp His Leu
            215                 220                 225
Tyr Leu Glu Thr Lys Thr Leu Gln Gly Thr Lys Gly Glu Asn Ser
            230                 235                 240
Leu Ser Ser Thr Gly Thr Phe Leu Val Asp Asn Ser Ser Val Asp
            245                 250                 255
Phe Gln Lys Phe Pro Asp Lys Glu Ile Leu Arg Met Ala Gly Pro
            260                 265                 270
Leu Thr Ala Asp Phe Ile Val Lys Ile Arg Asn Ser Gly Ser Ala
            275                 280                 285
Asp Ser Thr Val Gln Phe Ile Phe Tyr Gln Pro Ile Ile His Arg
            290                 295                 300
Trp Arg Glu Thr Asp Phe Phe Pro Cys Ser Ala Thr Cys Gly Gly
            305                 310                 315
Gly Tyr Gln Leu Thr Ser Ala Glu Cys Tyr Asp Leu Arg Ser Asn
            320                 325                 330
Arg Val Val Ala Asp Gln Tyr Cys His Tyr Tyr Pro Glu Asn Ile
            335                 340                 345
Lys Pro Lys Pro Lys Leu Gln Glu Cys Asn Leu Asp Pro Cys Pro
            350                 355                 360
Ala Ser Asp Gly Tyr Lys Gln Ile Met Pro Tyr Asp Leu Tyr His
            365                 370                 375
Pro Leu Pro Arg Trp Glu Ala Thr Pro Trp Thr Ala Cys Ser Ser
            380                 385                 390
Ser Cys Gly Gly Asp Ile Gln Ser Arg Ala Val Ser Cys Val Glu
            395                 400                 405
Glu Asp Ile Gln Gly His Val Thr Ser Val Glu Glu Trp Lys Cys
            410                 415                 420
Met Tyr Thr Pro Lys Met Pro Ile Ala Gln Pro Cys Asn Ile Phe
            425                 430                 435
```

```
Asp Cys Pro Lys Trp Leu Ala Gln Glu Trp Ser Pro Cys Thr Val
                440                 445                 450
Thr Cys Gly Gln Gly Leu Arg Tyr Arg Val Val Leu Cys Ile Asp
                455                 460                 465
His Arg Gly Met His Thr Gly Gly Cys Ser Pro Lys Thr Lys Pro
                470                 475                 480
His Ile Lys Glu Glu Cys Ile Val Pro Thr Pro Cys Tyr Lys Pro
                485                 490                 495
Lys Glu Lys Leu Pro Val Glu Ala Lys Leu Pro Trp Phe Lys Gln
                500                 505                 510
Ala Gln Glu Leu Glu Glu Gly Ala Ala Val Ser Glu Glu Pro Ser
                515                 520                 525
Phe Ile Pro Glu Ala Trp Ser Ala Cys Thr Val Thr Cys Gly Val
                530                 535                 540
Gly Thr Gln Val Arg Ile Val Arg Cys Gln Val Leu Leu Ser Phe
                545                 550                 555
Ser Gln Ser Val Ala Asp Leu Pro Ile Asp Glu Cys Glu Gly Pro
                560                 565                 570
Lys Pro Ala Ser Gln Arg Ala Cys Tyr Ala Gly Pro Cys Ser Gly
                575                 580                 585
Glu Ile Pro Glu Phe Asn Pro Asp Glu Thr Asp Gly Leu Phe Gly
                590                 595                 600
Gly Leu Gln Asp Phe Asp Glu Leu Tyr Asp Trp Glu Tyr Glu Gly
                605                 610                 615
Phe Thr Lys Cys Ser Glu Ser Cys Gly Gly Gly Val Gln Glu Ala
                620                 625                 630
Val Val Ser Cys Leu Asn Lys Gln Thr Arg Glu Pro Ala Glu Glu
                635                 640                 645
Asn Leu Cys Val Thr Ser Arg Arg Pro Pro Gln Leu Leu Lys Ser
                650                 655                 660
Cys Asn Leu Asp Pro Cys Pro Ala Ser Pro Val Ile
                665                 670
```

<210> 14
<211> 442
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1468733CD1

<400> 14

```
Met Val Glu Ala Met Glu Ala Met Met Ile Thr Met Ala Ile Met
  1                 5                  10                  15
Met Ala Met Asp Leu Gly Gln Ile Asp Leu Glu Glu Thr Ser Ile
                 20                  25                  30
Thr Val Phe Gln Glu Cys Leu Ile Thr Tyr Gly Asp Gly Gly Ser
                 35                  40                  45
Thr Phe Gln Ser Thr Thr Gly His Cys Val His Met Arg Gly Leu
                 50                  55                  60
Pro Tyr Arg Ala Thr Glu Asn Asp Ile Tyr Asn Phe Phe Ser Pro
                 65                  70                  75
Leu Asn Pro Val Arg Val His Ile Glu Ile Gly Pro Asp Gly Arg
                 80                  85                  90
Val Thr Gly Glu Ala Asp Val Glu Phe Ala Thr His Glu Asp Ala
                 95                 100                 105
Val Ala Ala Met Ser Lys Asp Lys Ala Asn Met Gln His Arg Tyr
                110                 115                 120
Val Glu Leu Phe Leu Asn Ser Thr Ala Gly Ala Ser Gly Gly Ala
                125                 130                 135
Tyr Glu His Arg Tyr Val Glu Leu Phe Leu Asn Ser Thr Ala Gly
                140                 145                 150
```

```
Ala Ser Gly Gly Ala Tyr Gly Ser Gln Met Met Gly Gly Met Gly
                155                 160                 165
Leu Ser Asn Gln Ser Ser Tyr Gly Gly Pro Ala Ser Gln Gln Leu
                170                 175                 180
Ser Gly Gly Tyr Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Leu
                185                 190                 195
Gly Gly Gly Leu Gly Asn Val Leu Gly Gly Leu Ile Ser Gly Ala
                200                 205                 210
Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
                215                 220                 225
Gly Gly Gly Gly Gly Thr Ala Met Arg Ile Leu Gly Gly Val Ile
                230                 235                 240
Ser Ala Ile Ser Glu Ala Ala Ala Gln Tyr Asn Pro Glu Pro Pro
                245                 250                 255
Pro Pro Arg Thr His Tyr Ser Asn Ile Glu Ala Asn Glu Ser Glu
                260                 265                 270
Glu Val Arg Gln Phe Arg Arg Leu Phe Ala Gln Leu Ala Gly Asp
                275                 280                 285
Asp Met Glu Val Ser Ala Thr Glu Leu Met Asn Ile Leu Asn Lys
                290                 295                 300
Val Val Thr Arg His Pro Asp Leu Lys Thr Asp Gly Phe Gly Ile
                305                 310                 315
Asp Thr Cys Arg Ser Met Val Ala Val Met Asp Ser Asp Thr Thr
                320                 325                 330
Gly Lys Leu Gly Phe Glu Glu Phe Lys Tyr Leu Trp Asn Asn Ile
                335                 340                 345
Lys Arg Trp Gln Ala Ile Tyr Lys Gln Phe Asp Thr Asp Arg Ser
                350                 355                 360
Gly Thr Ile Cys Ser Ser Glu Leu Pro Gly Ala Phe Glu Ala Ala
                365                 370                 375
Gly Phe His Leu Asn Glu His Leu Tyr Asn Met Ile Ile Arg Arg
                380                 385                 390
Tyr Ser Asp Glu Ser Gly Asn Met Asp Phe Asp Asn Phe Ile Ser
                395                 400                 405
Cys Leu Val Arg Leu Asp Ala Met Phe Arg Ala Phe Lys Ser Leu
                410                 415                 420
Asp Lys Asp Gly Thr Gly Gln Ile Gln Val Asn Ile Gln Glu Trp
                425                 430                 435
Leu Gln Leu Thr Met Tyr Ser
                440
```

<210> 15
<211> 378
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1652084CD1

<400> 15

```
Met Gly Ser Leu Ser Thr Ala Asn Val Glu Phe Cys Leu Asp Val
  1               5                  10                  15
Phe Lys Glu Leu Asn Ser Asn Asn Ile Gly Asp Asn Ile Phe Phe
                20                  25                  30
Ser Ser Leu Ser Leu Leu Tyr Ala Leu Ser Met Val Leu Leu Gly
                35                  40                  45
Ala Arg Gly Glu Thr Glu Glu Gln Leu Glu Lys Val Trp Asn Ser
                50                  55                  60
Ser Glu Val Leu His Phe Ser His Thr Val Asp Ser Leu Lys Pro
                65                  70                  75
Gly Phe Lys Asp Ser Pro Lys Pro Asp Ser Asn Cys Thr Leu Ser
                80                  85                  90
```

```
Ile Ala Asn Arg Leu Tyr Gly Thr Lys Thr Met Ala Phe His Gln
                 95                  100                 105
Gln Tyr Leu Ser Cys Ser Glu Lys Trp Tyr Gln Ala Arg Leu Gln
                110                 115                 120
Thr Val Asp Phe Glu Gln Ser Thr Glu Glu Thr Arg Lys Thr Ile
                125                 130                 135
Asn Ala Trp Val Glu Asn Lys Thr Asn Gly Lys Val Ala Asn Leu
                140                 145                 150
Phe Gly Lys Ser Thr Ile Asp Pro Ser Ser Val Met Val Leu Val
                155                 160                 165
Asn Ala Ile Tyr Phe Lys Gly Gln Trp Gln Asn Lys Phe Gln Val
                170                 175                 180
Arg Glu Thr Val Lys Ser Pro Phe Gln Leu Ser Glu Gly Lys Asn
                185                 190                 195
Val Thr Val Glu Met Met Tyr Gln Ile Gly Thr Phe Lys Leu Ala
                200                 205                 210
Phe Val Lys Glu Pro Gln Met Gln Val Leu Glu Leu Pro Tyr Val
                215                 220                 225
Asn Asn Lys Leu Ser Met Ile Ile Leu Leu Pro Val Gly Ile Ala
                230                 235                 240
Asn Leu Lys Gln Ile Glu Lys Gln Leu Asn Ser Gly Thr Phe His
                245                 250                 255
Glu Trp Thr Ser Ser Ser Asn Met Met Glu Arg Glu Val Glu Val
                260                 265                 270
His Leu Pro Arg Phe Lys Leu Glu Ile Lys Tyr Glu Leu Asn Ser
                275                 280                 285
Leu Leu Lys Pro Leu Gly Val Thr Asp Leu Phe Asn Gln Val Lys
                290                 295                 300
Ala Asp Leu Ser Gly Met Ser Pro Thr Lys Gly Leu Tyr Leu Ser
                305                 310                 315
Lys Ala Ile His Lys Ser Tyr Leu Asp Val Ser Glu Glu Gly Thr
                320                 325                 330
Glu Ala Ala Ala Ala Thr Gly Asp Ser Ile Ala Val Lys Ser Leu
                335                 340                 345
Pro Met Arg Ala Gln Phe Lys Ala Asn His Pro Phe Leu Phe Phe
                350                 355                 360
Ile Arg His Thr His Thr Asn Thr Ile Leu Phe Cys Gly Lys Leu
                365                 370                 375
Ala Ser Pro
```

<210> 16
<211> 458
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3456896CD1

<400> 16

```
Met Ala Pro Pro Ala Ala Arg Leu Ala Leu Leu Ser Ala Ala Ala
  1               5                  10                  15
Leu Thr Leu Ala Ala Arg Pro Ala Pro Ser Pro Gly Leu Gly Pro
                20                  25                  30
Gly Pro Glu Cys Phe Thr Ala Asn Gly Ala Asp Tyr Arg Gly Thr
                35                  40                  45
Gln Asn Trp Thr Ala Leu Gln Gly Gly Lys Pro Cys Leu Phe Trp
                50                  55                  60
Asn Glu Thr Phe Gln His Pro Tyr Asn Thr Leu Lys Tyr Pro Asn
                65                  70                  75
Gly Glu Gly Gly Leu Gly Glu His Asn Tyr Cys Arg Asn Pro Asp
                80                  85                  90
```

```
Gly Asp Val Ser Pro Trp Cys Tyr Val Ala Glu His Glu Asp Gly
                95                  100                 105
Val Tyr Trp Lys Tyr Cys Glu Ile Pro Ala Cys Gln Met Pro Gly
            110                 115                 120
Asn Leu Gly Cys Tyr Lys Asp His Gly Asn Pro Pro Pro Leu Thr
        125                 130                 135
Gly Thr Ser Lys Thr Ser Asn Lys Leu Thr Ile Gln Thr Cys Ile
    140                 145                 150
Ser Phe Cys Arg Ser Gln Arg Phe Lys Phe Ala Gly Met Glu Ser
155                 160                 165
Gly Tyr Ala Cys Phe Cys Gly Asn Asn Pro Asp Tyr Trp Lys Tyr
            170                 175                 180
Gly Glu Ala Ala Ser Thr Glu Cys Asn Ser Val Cys Phe Gly Asp
        185                 190                 195
His Thr Gln Pro Cys Gly Gly Asp Gly Arg Ile Ile Leu Phe Asp
    200                 205                 210
Thr Leu Val Gly Ala Cys Gly Gly Asn Tyr Ser Ala Met Ser Ser
215                 220                 225
Val Val Tyr Ser Pro Asp Phe Pro Asp Thr Tyr Ala Thr Gly Arg
            230                 235                 240
Val Cys Tyr Trp Thr Ile Arg Val Pro Gly Ala Ser His Ile His
        245                 250                 255
Phe Ser Phe Pro Leu Phe Asp Ile Arg Asp Ser Ala Asp Met Val
    260                 265                 270
Glu Leu Leu Asp Gly Tyr Thr His Arg Val Leu Ala Arg Phe His
275                 280                 285
Gly Arg Ser Arg Pro Pro Leu Ser Phe Asn Val Ser Leu Asp Phe
            290                 295                 300
Val Ile Leu Tyr Phe Phe Ser Asp Arg Ile Asn Gln Ala Gln Gly
        305                 310                 315
Phe Ala Val Leu Tyr Gln Ala Val Lys Glu Glu Leu Pro Gln Glu
    320                 325                 330
Arg Pro Ala Val Asn Gln Thr Val Ala Glu Val Ile Thr Glu Gln
335                 340                 345
Ala Asn Leu Ser Val Ser Ala Ala Arg Ser Ser Lys Val Leu Tyr
            350                 355                 360
Val Ile Thr Thr Ser Pro Ser His Pro Pro Gln Thr Val Pro Gly
        365                 370                 375
Trp Thr Val Tyr Gly Leu Ala Thr Leu Leu Ile Leu Thr Val Thr
    380                 385                 390
Ala Ile Val Ala Lys Ile Leu Leu His Val Thr Phe Lys Ser His
395                 400                 405
Arg Val Pro Ala Ser Gly Asp Leu Arg Asp Cys His Gln Pro Gly
            410                 415                 420
Thr Ser Gly Glu Ile Trp Ser Ile Phe Tyr Lys Pro Ser Thr Ser
        425                 430                 435
Ile Ser Ile Phe Lys Lys Lys Leu Lys Gly Gln Ser Gln Gln Asp
    440                 445                 450
Asp Arg Asn Pro Leu Val Ser Asp
455
```

<210> 17
<211> 993
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7482256CB1

<400> 17
atgggcgcgc gcggggcgct gctgctggcg ctgctgctgg ctcgggctgg actcgggaag 60
ccggaggcct gcggccaccg ggaaattcac gcgctggtgg cgggcggagt ggagtccgcg 120

cgcgggcgct ggccatggca ggccagcctg cgcctgagga gacgccaccg atgtggaggg 180
agcctgctca gccgccgctg ggtgctctcg gctgcgcact gcttccaaaa cagtcgttac 240
aaagtgcagg acatcattgt gaaccctgac gcacttgggg ttttacgcaa tgacattgcc 300
ctgctgagac tggcctcttc tgtcacctac aatgcgtaca tccagcccat ttgcatcgag 360
tcttccacct tcaacttcgt gcaccggccg gactgctggg tgaccggctg ggggttaatc 420
agccccagtg gcacacctct gccacctcct tacaacctcc gggaagcaca ggtcaccatc 480
ttaaacaaca ccaggtgtaa ttacctgttt gaacagccct ctagccgtag tatgatctgg 540
gattccatgt tttgtgctgg tgctgaggat ggcagtgtag acacctgcaa aggtgactca 600
ggtggaccct tggtctgtga caaggatgga ctgtggtatc aggttggaat cgtgagctgg 660
ggaatggact gcggtcaacc caatcggcct ggtgtctaca ccaacatcag tgtgtacttc 720
cactggatcc ggagggtgat gtcccacagt acaccaaggc caaaccctcc ccagctgttg 780
ctgctccttg ccctgctgtg ggctccctga ctcctgcagc cattctgagt gcaccagaaa 840
ctgtgaggct gcagtgggga ccacagtatt ggctcacctc ctctgggctg tgggcgcttc 900
agggacaggg ttgggactgc ctgctggatc agattccggc ccctttgtc tcgtttgcta 960
ataaatacgt gtgcatgttc aaaaaaaaaa aaa 993

<210> 18
<211> 1238
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71973513CB1

<400> 18
atgaggggcc ttgtggtatt ccttgcagtc tttgctctct ctgaggtcaa tgccatcacc 60
agggttcctc tgcacaaagg gaagtcgctg aggagggccc tgaaggagcg caggctcctg 120
gaggacttcc tgaggaatca ccattatgca gtcagcagga agcactccag ctctggggtg 180
gtggccagcg agtctctgac caactacctg gattgtcagt actttgggaa gatctacatc 240
gggacccttc cccagaagtt caccttggtg tttgatacag gctccccgga tatctgggtg 300
ccctctgtct actgcaacag tgatgcctgt cagaaccacc aacgcttcga tccgtccaag 360
tcctccaccc agaacatggg caagtccctg tccatccagt atggcacagg cagcatgcgg 420
ggcttgctgg gctatgacac tgtcaccgtc tccaacattg tggacccca ccagactgtg 480
ggtctgagca cccaggaacc tggcgacgtc ttcacctact ccgagtttga tgggatcctg 540
gggctggcct atccctctct tgcctctgag tacgcgctgc gccttggttt caggaatgac 600
caggggagca tgctcacgct gagggccatt gatctgtcgt actacacagg ctccctgcac 660
tggataccca tgactgcaag aatactggca gttcactgtg gacaggaagg acctggggag 720
ggagggctgg atgaggccat cttgcatacc tttggaagtg tcatcattga cggcgtggtg 780
gtggcctgtg acggtggctg tcaggccatc ctggacaccg gcacctccct gctggtgggg 840
cctggtggca acatcctcaa catccagcag gccattggac gcactgcggg ccagtacaat 900
gagtttgaca tcgactgcgg gcgcctgagc agcattccca cggctgtctt cgagatccac 960
ggcaagaagt acccctgcc accctccgcc tataccagcc aggaccaggg cttctgcacc 1020
agtggtttcc agggtgacta tagttcccag cagtggatcc tggggaatgt cttcatctgg 1080
gagtattaca gtgtctttga caggaccaat aaccgtgtgg ggctggcgaa ggctgtctga 1140
ttgcatcact ggccacggac ctcaatgtga ccaaacacac acgcgcacat agatgagatg 1200
tgcaggcaga tggttcccaa taaacaccgc atttctgc 1238

<210> 19
<211> 1233
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7648238CB1

<400> 19
gggaagtatg acgtccaggg tccaagggca gccctgatgc tcagcagccc tggggtggcg 60
gccgctgtag tcactgccct ggaggacgtg ttccaggccc tgggctttga gagctgcgag 120
aggagggagg tcccggtcca gggcttcctc gaggaactgg cttggttcca ggagcagctg 180
gatgcccacg ggcgccctgt gggagggcag ctgaggcagc cacagcagct ggtccgggag 240
ctgagcggct gccgggccct gcggggctgc cccaaagtct tcctgctgct ctcaagtggt 300

cctgggtcct ccctggagcc cggagccttc cttgctggcc tgagagagct gtgtggccgc 360
tctcctcact ggtccctggt gcagctgctg acgaagctct tccgcagggt ggctgaagag 420
tccgcagggg gcacctgctg ccccgtcctt cggagctcct tgaggggggc actgtgcctg 480
ggaggcgtgg agccctggag gcctgagccg gcccccggtc ccagcacaca gtatgacctg 540
tccaaggcca gggctgccct cctcctggct gtgatccaag gccggcctgg ggcccagcat 600
gacgtggagg cgctgggggg cctgtgctgg gccctgggct ttgagaccac cgtgagaacg 660
gaccctacag cccaggcttt ccaggaggag ctggcccagt tccgggagca actggacacc 720
tgcaggggcc ctgtgagctg tgcccttgtg gccctgatgg cccatggggg accacggggt 780
cagctgctgg gggctgacgg gcaagaggtg cagcccgagg cactcatgca ggagctgagc 840
cgctgccagg tgctgcaggg ccgccccaag atcttcctgt tgcaggcctg ccgtggggga 900
aacagggatg ctggtgtggg gcccacagct ctcccctggt actggagctg gctgcgggca 960
cctccatctg tcccctccca tgcagatgtc ctgcagatct acgctgaggc ccaaggctat 1020
gtggcctatc gcgatgacaa gggctcagac tttatccaga cactggtgga ggtcctcaga 1080
gccaaccccg ggagagacct tctggagctg ctgactgagg tcaacaggcg ggtgtgcgag 1140
caggaggtgc tgggccccga ctgcgatgaa ctccgcaagg cctgcctgga gatccgcagc 1200
tcgctccggc gccggctctg cctccaggcc tga 1233

<210> 20
<211> 5511
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1719204CB1

<400> 20
atggctccac tccgcgcgct gctgtcctac ctgctgcctt tgcactgtgc gctctgcgcc 60
gccgcgggca gccggacccc agagctgcac ctctctggaa agctcagtga ctatggtgtg 120
acagtgccct gcagcacaga ctttcgggga cgcttcctct cccacgtggt gtctggccca 180
gcagcagcct ctgcagggag catggtagtg gacacgccac ccacactacc acgacactcc 240
agtcacctcc gggtggctcg cagccctctg cacccaggag ggaccctgtg gcctggcagg 300
gtggggcgcc actccctcta cttcaatgtc actgttttcg ggaaggaact gcacttgcgc 360
ctgcggccca atcggaggtt ggtagtgcca ggatcctcag tggagtggca ggaggatttt 420
cgggagctgt tccggcagcc cttacggcag gagtgtgtgt acactggagg tgtcactgga 480
atgcctgggg cagctgttgc catcagcaac tgtgacggat tggcgggcct catccgcaca 540
gacagcaccg acttcttcat tgagcctctg gagcggggcc agcaggagaa ggaggccagc 600
gggaggacac atgtggtgta ccgccgggag gccgtccagc aggagtgggc agaacctgac 660
ggggacctgc acaatgaagc ctttggcctg ggagaccttc ccaacctgct gggcctggtg 720
ggggaccagc tgggcgacac agagcggaag cggcggcatg ccaagccagg cagctacagc 780
atcgaggtgc tgctggtggt ggacgactcg gtggttcgct tccatggcaa ggagcatgtg 840
cagaactatg tcctcaccct catgaatatc gtagatgaga tttaccacga tgagtccctg 900
ggggttcata taaatattgc cctcgtccgc ttgatcatgg ttggctaccg acagtccctg 960
agcctgatcg agcgcgggaa cccctcacgc agcctggagc aggtgtgtcg ctgggcacac 1020
tcccagcagc gccaggaccc cagccacgct gagcaccatg accacgttgt gttcctcacc 1080
cggcaggact ttgggccctc agggtatgca cccgtcactg gcatgtgtca ccccctgagg 1140
agctgtgccc tcaaccatga ggatggcttc tcctcagcct tcgtgatagc tcatgagacc 1200
ggccacgtgc tcggcatgga gcatgacggt caggggaatg gctgtgcaga tgagaccagc 1260
ctgggcagcg tcatggcgcc cctggtgcag gctgccttcc accgcttcca ttggtcccgc 1320
tgcagcaagc tggagctcag ccgctacctc ccctcctacg actgcctcct cgatgacccc 1380
tttgatcctg cctggcccca gcccccagag ctgcctggga tcaactactc aatggatgag 1440
cagtgccgct ttgactttgg cagtggctac cagacctgct tggcattcag gacctttgag 1500
ccctgcaagc agctgtggtg cagccatcct gacaacccgt acttctgcaa gaccaagaag 1560
gggcccccgc tggatgggac tgagtgtgca cccggcaagt ggtgcttcaa aggtcactgc 1620
atctggaagt cgccggagca gacatatggc caggatggag gctggagctc ctggaccaag 1680
tttgggtcat gttcgcggtc atgtgggggc ggggtgcgat cccgcagccg gagctgcaac 1740
aacccctccc tatggagccg cccgtgctta gggcccatgt tcgagtacca ggtctgcaac 1800
agcgaggagt gccctgggac ctacgaggac ttccgggccc agcagtgtgc caagcgcaac 1860
tcgtactatg tgcaccagaa tgccaagcac agctgggtgc cctacgagcc tgacgatgac 1920
gcccagaagt gtgagctgat ctgccagtcg gcggacacgg gggacgtggt gttcatgaac 1980
caggtggttc acgatgggac acgctgcagc taccgggacc catacagcgt ctgtgcgcgt 2040
ggcgagtgtg tgcctgtcgg ctgtgacaag gaggtggggt ccatgaaggc ggatgacaag 2100
tgtggagtct gcgggggtga caactcccac tgcaggactg tgaaggggac gctgggcaag 2160

gcctccaagc aggcaggagc tctcaagctg gtgcagatcc cagcaggtgc caggcacatc 2220
cagattgagg cactggagaa gtcccccac cggtcagtgg tgaagaacca ggtcaccggc 2280
agcttcatcc tcaaccccaa gggcaaggaa gccacaagcc ggaccttcac cgccatgggc 2340
ctggagtggg aggatgcggt ggaggatgcc aaggaaagcc tcaagaccag cgggcccctg 2400
cctgaagcca ttgccatcct ggctctcccc ccaactgagg gtggcccccg cagcagcctg 2460
gcctacaagt acgtcatcca tgaggacctg ctgcccctta tcgggagcaa caatgtgctc 2520
ctggaggaga tggacaccta tgagtgggcg ctcaagagct gggcccctg cagcaaggcc 2580
tgtggaggag ggatccagtt caccaaatac ggctgccggc gcagacgaga ccaccacatg 2640
gtgcagcgac acctgtgtga ccacaagaag aggcccaagc ccatccgccg gcgctgcaac 2700
cagcacccgt gctctcagcc tgtgtgggtg acggaggagt ggggtgcctg cagccggagc 2760
tgtgggaagc tgggggtgca gacacggggg atacagtgcc tgctgcccct ctccaatgga 2820
acccacaagg tcatgccggc caaagcctgc gccggggacc ggcctgaggc ccgacggccc 2880
tgtctccgag tgccctgccc agcccagtgg aggctgggag cctggtccca gtgctctgcc 2940
acctgtggag agggcatcca gcagcggcag gtggtgtgca ggaccaacgc caacagcctc 3000
gggcattgcg agggggatag gccagacact gtccaggtct gcagcctgcc cgcctgtgga 3060
ggaaatcacc agaactccac ggtgagggcc gatgtctggg aacttgggac gccagagggg 3120
cagtgggtgc cacaatctga acccctacat cccattaaca agatatcatc aacggagccc 3180
tgcacgggag acaggtctgt cttctgccag atggaagtgc tcgatcgcta ctgctccatt 3240
cccggctacc accggctctg ctgtgtgtcc tgcatcaaga aggcctcggg ccccaaccct 3300
ggcccagacc ctggcccaac ctcactgccc cccttctcca ctcctggaag ccccttacca 3360
ggaccccagg accctgcaga tgctgcagag cctcctggaa agccaacggg atcagaggac 3420
catcagcatg gccgagccac acagctccca ggagctctgg atacaagctc cccagggacc 3480
cagcatccct ttgcccctga gacaccaatc cctggagcat cctggagcat ctcccctacc 3540
acccccgggg ggctgccttg gggctggact cagacaccta cgccagtccc tgaggacaaa 3600
gggcaacctg gagaagacct gaggcatccc ggcaccagcc tccctgctgc ctccccggtg 3660
acatgagctg tgccctgcca tcccactggc acgtttacac tctgtgtact gccccgtgac 3720
tcccagctca gaggacacac atagcagggc aggcgcaagc acagacttca ttttaaatca 3780
ttcgccttct tctcgtttgg ggctgtgatg ctctttaccc cacaaagcgg ggtgggagga 3840
agacaaagat cagggaaagc cctaatcgga gatacctcag caagctgccc ccggcgggac 3900
tgaccctctc agggcccctg ttggtctccc ctgccaagac cagggtcaac tattgctccc 3960
tcctcacaga ccctgggcct gggcagatct gaatcccggc tggtctgtag ctagaagctg 4020
tcagggctgc ctgccttccc ggaactgtga ggacccctgt ggaggccctg catatttggc 4080
ccctctcccc agaaaggcaa agcagggcca gggtaggtgg gggactgttc acagccaggc 4140
cgagaggagg ggggcctggg aatgtggcat gaggcttccc agctgcaggg ctggaggggg 4200
tggaacacaa gatgatcgca ggcccagctc ctggaagcca agagctccat gcagttccac 4260
cagctgaggc caggcagcag aggccagttt gtctttgctg gccagaagat ggtgctcatg 4320
gccatactct ggccttgcag atgtcactag tgttacttct agtgactcca gattacagac 4380
tggcccccca atctcacccc agcccaccag agaagggggc tcaggacacc ctggacccca 4440
agtcctcagc atccagggat ttccaaactg gcgctcaccc cctgactcca ccaggatggc 4500
aacttcaatt atcactctca gcctggaagg ggactctgtg ggacacagag ggaacacgat 4560
ttctcaggct gtcccttcaa tcattgccct tctccgaaga tcgctcctgc tggagtcgga 4620
catcttcatc ttctacctgg ctcaagctgg gccagagtgt gtggttctcc caggggtggt 4680
tggaccccag gactgaggac cagagtccac tcatagcctg gccctggaga tgacaagggc 4740
cacccaggcc aagtgcccca gggcagggtg ccagcccctg gcctggtgct ggagtgggga 4800
agacacactc acccacggtg ctgtaagggc ctgagctgtg ctcagctgcc ggccatgcta 4860
cctccaaggg acaggtaaca gtcttagatc ctctggctct caggaagtgg cagggggtcc 4920
caggacacct ccggggtctt ggaggatgtc tcctaaactc ctgccaggtg atagaggtgc 4980
ttctcacttc ttccttcccc aaggcaaagg ggctgttctg agccagcctg gaggaacatg 5040
agtagtgggc ccctggcctg caaccccttt ggagagtgga ggtcctgggg ggctccccgc 5100
cctccccctg ttgccctccc ctccctggga tgctggggca cacgtggagt cattcctgtg 5160
agaaccagcc tggcctgtgt taaactcttg tgccttggaa atccagatct ttaaaatttt 5220
atgtatttat taacatcgcc attgggcccc aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 5280
aaaaaaaaaa aaaaaaaaaa aaaggggggg ggcccgcaaa aaggggccc cgacaccgcg 5340
ggaaaataaa ccggcgccgg accccggggg ggggtggacc aattgagcct aacacacgag 5400
gggggggtgc ccggttttgt aaaaacaccc gggggaaatg tgacccgcac actatagggg 5460
cgccgcagag gggcccaaac caggcacggg gcggaggaga aacggagccc g 5511

<210> 21
<211> 7142
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 7472647CB1

<400> 21
aatgtgagag gggctgatgg aagctgatag gcaggactgg agtgttagca ccagtactgg 60
atgtgacagc aggcagagga gcacttagca gcttattcag tgtccgattc tgattccggc 120
aaggatccaa gcatggaatg ctgccgtcgg gcaactcctg gcacactgct cctctttctg 180
gctttcctgc tcctgagttc caggaccgca cgctccgagg aggaccggga cggcctatgg 240
gatgcctggg gcccatggag tgaatgctca cgcacctgcg ggggaggggc ctcctactct 300
ctgaggcgct gcctgagcag caagagctgt gaaggaagaa atatccgata cagaacatgc 360
agtaatgtgg actgcccacc agaagcaggt gatttccgag ctcagcaatg ctcagctcat 420
aatgatgtca agcaccatgg ccagttttat gaatggcttc ctgtgtctaa tgaccctgac 480
aacccatgtt cactcaagtg ccaagccaaa ggaacaaccc tggttgttga actagcacct 540
aaggtcttag atggtacgcg ttgctataca gaatctttgg atatgtgcat cagtggttta 600
tgccaaattg ttggctgcga tcaccagctg ggaagcaccg tcaaggaaga taactgtggg 660
gtctgcaacg gagatgggtc cacctgccgg ctggtccgag ggcagtataa atcccagctc 720
tccgcaacca aatcggatga tactgtggtt gcaattccct atggaagtag acatattcgc 780
cttgtcttaa aaggtcctga tcacttatat ctggaaacca aaaccctcca ggggactaaa 840
ggtgaaaaca gtctcagctc cacaggaact ttccttgtgg acaattctag tgtggacttc 900
cagaaatttc cagacaaaga gatactgaga atggctggac cactcacagc agatttcatt 960
gtcaagattc gtaactcggg ctccgctgac agtacagtcc agttcatctt ctatcaaccc 1020
atcatccacc gatggaggga gacggatttc tttccttgct cagcaacctg tggaggaggt 1080
tatcagctga catcggctga gtgctacgat ctgaggagca accgtgtggt tgctgaccaa 1140
tactgtcact attacccaga gaacatcaaa cccaaaccca agcttcagga gtgcaacttg 1200
gatccttgtc cagccagtga cggatacaag cagatcatgc cttatgacct ctaccatccc 1260
cttcctcggt gggaggccac cccatggacc gcgtgctcct cctcgtgtgg gggggacatc 1320
cagagccggg cagtttcctg tgtggaggag gacatccagg ggcatgtcac ttcagtggaa 1380
gagtggaaat gcatgtacac ccctaagatg cccatcgcgc agccctgcaa catttttgac 1440
tgccctaaat ggctggcaca ggagtggtct ccgtgcacag tgacgtgtgg ccagggcctc 1500
agataccgtg tggtcctctg catcgaccat cgaggaatgc acacaggagg ctgtagccca 1560
aaaacaaagc cccacataaa agaggaatgc atcgtaccca ctccctgcta taaacccaaa 1620
gagaaacttc cagtcgaggc caagttgcca tggttcaaac aagctcaaga gctagaagaa 1680
ggagctgctg tgtcagagga gccctcgttc atcccagagg cctggtcggc ctgcacagtc 1740
acctgtggtg tggggaccca ggtgcgaata gtcaggtgcc aggtgctcct gtctttctct 1800
cagtccgtgg ctgacctgcc tattgacgag tgtgaagggc ccaagccagc atcccagcgt 1860
gcctgttatg caggcccatg cagcggggaa attcctgagt tcaacccaga cgagacagat 1920
gggctctttg gtggcctgca ggatttcgac gagctgtatg actgggagta tgaggggttc 1980
accaagtgct ccgagtcctg tggaggaggg cccgggcggc catccacgaa gcacagcccg 2040
cacatcgcgg ccgccaggaa ggtctacatc cagactcgca ggcagaggaa gctgcacttc 2100
gtggtggggg gcttcgccta cctgctcccc aagacggcgg tggtgctgcg ctgcccggcg 2160
cgcagggtcc gcaagcccct catcacctgg gagaaggacg gccagcacct catcagctcg 2220
acgcacgtca cggtggcccc cttcggctat ctcaagatcc accgcctcaa gccctcggat 2280
gcaggcgtct acacctgctc agcgggcccg gcccgggagc actttgtgat taagctcatc 2340
ggaggcaacc gcaagctcgt ggcccggccc ttgagcccga gaagtgagga agaggtgctt 2400
gcggggagga agggcggccc gaaggaggcc ctgcagaccc acaaacacca gaacgggatc 2460
ttctccaacg gcagcaaggc ggagaagcgg ggcctggccg ccaacccggg gagccgctac 2520
gacgacctcg tctcccggct gctggagcag ggcggctggc ccggagagct gctggcctcg 2580
tgggaggcgc aggactccgc ggaaaggaac acgacctcgg aggaggaccc gggtgcagag 2640
caagtgctcc tgcacctgcc cttcaccatg gtgaccgagc agcggcgcct ggacgacatc 2700
ctggggaacc tctcccagca gcccgaggag ctgcgcgacc tctacagcaa gcacctggtg 2760
gcccagctgg cccaggagat cttccgcagc cacctggagc accaggacac gctcctgaag 2820
ccctcggagc gcaggacttc cccagtgact ctctcgcctc ataaacacgt gtctggcttc 2880
agcagctccc tgcggacctc ctccaccggg gacgccgggg gaggctctcg aaggccacac 2940
cgcaagccca ccatcctgcg caagatctca gcggcccagc agctctcagc ctcggaggtg 3000
gtcacccacc tggggcagac ggtggccctg gccagcggga cactgagtgt tcttctgcac 3060
tgtgaggcca tcggccaccc aaggcctacc atcagctggg ccaggaatgg agaagaagtt 3120
cagttcagtg acaggattct tctacagcca gatgattcct tacagatctt ggcaccagtg 3180
gaagcagatg tgggtttcta cacttgcaat gccaccaatg ccttgggata cgactctgtc 3240
tccattgccg tcacattagc aggaaagcca ctagtgaaaa cgtcacgaat gacagtgatc 3300
aacacggaga agcctgcagt cacagtcgat ataggaagca ccatcaaaac agtgcaggga 3360
gtgaatgtga caatcaactg ccaggttgca ggagtgcctg aagctgaagt cacttggttc 3420
aggaataaaa gcaaactggg ctccccgcac catctgcacg aaggctcctt gctgctcaca 3480
aacgtgtcct cctcggatca gggcctgtac tcctgcaggg cggccaatct tcatggagag 3540

ctgactgaga gcacccagct gctgatccta gatccccccc aagtccccac acagttggaa 3600
gacatcaggg ccttgctcgc tgccactgga ccgaaccttc cttcagtgct gacgtctcct 3660
ctgggaacac agctggtcct gggtcctggg aattctgctc tccttggctg ccccatcaaa 3720
ggtcaccctg tccctaatat cacctggttt catggtggtc agccaattgt cactgccaca 3780
ggactgacgc atcacatctt ggcagctgga cagatccttc aagttgcaaa ccttagcggt 3840
gggtctcaag gggaattcag ctgccttgct cagaatgagg caggggtgct catgcagaag 3900
gcatctttag tgatccaaga ttactggtgg tctgtggaca gactggcaac ctgctcagcc 3960
tcctgtggta accgggggggt tcagcagccc cgcttgaggt gcctgctgaa cagcacggag 4020
gtcaaccctg cccactgcgc agggaaggtt cgccctgcgg tgcagcccat cgcgtgcaac 4080
cggagagact gcccttctcg gtggatggtg acctcctggt ctgcctgtac ccggagctgt 4140
gggggaggtg tccagacccg cagggtgacc tgtcaaaagc tgaaagcctc tgggatctcc 4200
acccctgtgt ccaatgacat gtgcacccag gtcgccaagc ggcctgtgga cacccaggcc 4260
tgtaaccagc agctgtgtgt ggagtgggcc ttctccagct ggggccagtg caatgggcct 4320
tgcatcgggc ctcacctagc tgtgcaacac agacaagtct tctgccagac acgggatggc 4380
atcaccttac catcagagca gtgcagtgct cttccgaggc ctgtgagcac ccagaactgc 4440
tggtcagagg cctgcagtgt acactggaga gtcagcctgt ggaccctgtg cacagctacc 4500
tgtggcaact acggcttcca gtcccggcgt gtggagtgtg tgcatgcccg caccaacaag 4560
gcagtgcctg agcacctgtg ctcctggggg ccccggcctg ccaactggca gcgctgcaac 4620
atcaccccat gtgaaaacat ggagtgcaga gacaccacca ggtactgcga gaaggtgaaa 4680
cagctgaaac tctgccaact cagccagttt aaatctcgct gctgtggaac ttgtggcaaa 4740
gcgtgaagat agggtgtggg gaaaaactct accctggcca cacgaaggac tcacgcaacc 4800
acctcggaca gaacctaagc tttcttcatt ttatttattt atttcccccc ccccactcca 4860
cacacaccct tccaacctcc tccacctcca ccttcaagca taaggacgtc cgcgtgtttt 4920
ctctttcagt tagctggagg acaggatgtt gggaaaggaa aggacagatg tctaaaggag 4980
gttgcagagc aggccaggca gacagtgggg gctcccttga agagcttcct ccctcccaaa 5040
cctgggtctc aaagacctag aaagaggcag gcacagcccc tgcggacagc agggagccag 5100
aaggtttgta gcctattggt gcaaacattg gacaaattcc tgtgtctttc ctagaagcgc 5160
actatcacaa acacaggagt gttttgctcc tttgtctcct cttccccatc tatgtccctt 5220
tagtcacagt taggacaaat ggggagggga caccatgctg aggcagaaac tagcccagaa 5280
ctcactcagt tcttctagtg ggtgagtgca gagagagaag aactcagatc accagtaggg 5340
agaggtaaaa aagcaaacaa agcaggctct aaggcacaca acattgcaga aaatgaggaa 5400
gggaggggag ggaagggaca gaagcaaaaa ggagcctgtg gtgttcccca gtggggcagg 5460
gtgagcaggg gcttccaggc tgcatgaggc tcatggacca gctctgatcc catgcatgtg 5520
cgcatgctca gagccctgct gcccacaaca gagcactgcg ctgcgtggga gtccccactt 5580
cccaagctat cagagtcaac gtcctgcctg tgcagctgca gcaaagccag tgagaggtgg 5640
gtctcgccat gcagtaaggc caccctggca cctctttatc taaatccgaa gtcccctagc 5700
cccgcactaa ctaactgctg ctgtgggcca gggccatttt gagcatgaat ggcccaggtt 5760
ttttgccttc taggaccttt gctgctccac cgaagggcca gggactatgg ttaacttatc 5820
aacatcaacc cattaactag tcactgtgcc agagagtatc tgtcaggctg tcaggttgta 5880
gcaacctctt cattccagag ctggcccagg gaccggggtg ggacaatggg tttatgcgtg 5940
tccacagtac accctcccctc tcccagcctc caccccaggg tctgcaggtc ctccggcatg 6000
tagtatttat ctagcaaggc ggggtggtgg aggcagcacc ctggcaaagc agctcacaca 6060
ctgcagccac actcatcagc tgtggtgagg cggctggagc aaagtcaaag tcatgcagca 6120
aaatgaaaac tctgggactc ttcggcaaaa tcctcattaa gccgagcagc tttggccaag 6180
taatttttgc ctccttccct cgcgtggcct gagtttagga gcaagggtgg ccagagtccc 6240
ttacccacag ataagcctcc cctcatgaaa tgccactcac cccgggctac cattgacatc 6300
agggctgcat ttccagccag cctggaagta aaatttgaga ggaagacaat attaatctgt 6360
gtccccacct agtgagctgt ggacaggttt aagttgggtc tccttcttct tcaccacaaa 6420
aacaggctct aagaaatcat gttactaaaa aatcagtgta aagtctgttt aaaataaaaa 6480
agaatgtttt ctatgtctgt atatcttttg tgaatattta ttaggatttc ttattaaaaa 6540
agtgcaatat taataattgt acattgtcat ccagaaacaa aactattggg gggactttat 6600
taactaactt cctgcagttg tgttcctgta aactcagtag tgattattat atttttccta 6660
tttttaatag aacctggtgt ttaactctgg atccattcac tgtacaggat gtgttgtaaa 6720
aactaacatg ggatgctgag gcagtaagag ggaattcatt tgtggcataa tagttatgca 6780
tggaatgata aagacagaca aattccatac tactactaat gtggttaatt atttctagtt 6840
cgatagtgat tgaaaatcag tggtcactat ttacatttcc taaagagcaa gcatcctcca 6900
gctccatgtt gggttggagc agttggcagt gggtctcagt gagctggcag aacctaggtt 6960
tgggtgggaa gcagaatgct cgttgcatga aatgaatgta catttaatgt ttgttctgtg 7020
aattgcaact cagcagcacc acaagacaat gaaggctgct ggctaatgtg gaaggaggca 7080
ctttctcctc taaaacacaa aactgtattt gtatttttg tacagataat acagcttatc 7140
ta 7142

<210> 22

<211> 6565
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472654CB1

<400> 22
aagttttaaa gaaataaaat tgttatgctt cgattttggt atggtattga ctctttagca 60
cataggtagc cctcaaaaaa atcatccagt tttctaaatt atggaaattt tgtggaagac 120
gttgacctgg attttgagcc tcatcatggc ttcatcggaa tttcatagtg accacaggct 180
ttcatacagt tctcaagagg aattcctgac ttatcttgaa cactaccagc taactattcc 240
aataagggtt gatcaaaatg gagcatttct cagctttact gtgaaaaatg ataaacactc 300
aaggagaaga cggagtatgg accctattga tccacagcag gcagtatcta agttattttt 360
taaactttca gcctatggca agcactttca tctaaacttg actctcaaca cagattttgt 420
gtccaaacat tttacagtag aatattgggg gaaagatgga ccccagtgga aacatgattt 480
tttagacaac tgtcattaca caggatattt gcaagatcaa cgtagtacaa ctaaagtggc 540
tttaagcaac tgtgttgggt tgcatggtgt tattgctaca gaagatgaag agtattttat 600
cgaaccttta aagaatacca cagaggattc caagcatttt agttatgaaa atggccaccc 660
tcatgttatt tacaaaaagt ctgcccttca acaacgacat ctgtatgatc actctcattg 720
tggggtttcg gatttcacaa gaagtggcaa accttggtgg ctgaatgaca catccactgt 780
ttcttattca ctaccgatta acaacacaca tatccaccac agacagaaga gatcagtgag 840
cattgaacgg tttgtggaga cattggtagt ggcagacaaa atgatggtgg gctaccatgg 900
ccgcaaagac attgaacatt acattttgag tgtgatgaat attgttgcca aactttaccg 960
tgattccagc ctaggaaacg ttgtgaatat tatagtggcc cgcttaattg ttctcacaga 1020
agatcagcca aacttggaga taaaccacca tgcagacaag tccctcgata gcttctgtaa 1080
atggcagaaa tccattctct cccaccaaag tgatggaaac accattccag aaaatgggat 1140
tgcccaccac gataatgcag ttcttattac tagatatgat atctgcactt ataaaaataa 1200
gccctgtgga acactgggct tggcctctgt ggctggaatg tgtgagcctg aaaggagctg 1260
cagcattaat gaagacattg gcctgggttc agcttttacc attgcacatg agattggtca 1320
caattttggt atgaaccatg atggaattgg aaattcttgt gggacgaaag gtcatgaagc 1380
agcaaaactt atggcagctc acattactgc gaataccaat ccttttttcct ggtctgcttg 1440
cagtcgagac tacatcacca gctttctaga ttcaggccgt ggtacttgcc ttgataatga 1500
gcctcccaag cgtgactttc tttatccagc tgtggcccca ggtcaggtgt atgatgctga 1560
tgagcaatgt cgtttccagt atggagcaac ctcccgccaa tgtaaatatg gggaagtgtg 1620
tagagagctc tggtgtctca gcaaaagcaa ccgctgtgtc accaacagta ttccagcagc 1680
tgaggggaca ctgtgtcaaa ctgggaatat tgaaaaaggg tggtgttatc agggagattg 1740
tgttcctttt ggcacttggc cccagagcat agatgggggc tggggtccct ggtcactatg 1800
gggagagtgc agcaggacct gcgggggagg cgtctcctca tccctaagac actgtgacag 1860
tccagctttt ttcagacctt caggaggtgg aaaatattgc cttggggaaa ggaaacggta 1920
tcgctcctgt aacacagatc catgcccttt gggttcccga gattttcgag agaaacagtg 1980
tgcagacttt gacaatatgc ctttccgagg aaagtattat aactggaaac cctatactgg 2040
aggtggggta aaaccttgtg cattaaactg cttggctgaa ggttataatt tctacactga 2100
acgtgctcct gcggtgatcg atgggaccca gtgcaatgcg gattcactgg atatctgcat 2160
caatggagaa tgcaagcacg taggctgtga taatattttg ggatctgatg ctagggaaga 2220
tagatgtcga gtctgtggag gggacggaag cacatgtgat gccattgaag ggttcttcaa 2280
tgattcactg cccaggggag gctacatgga agtggtgcag ataccaagag gctctgttca 2340
cattgaagtt agagaagttg ccatgtcaaa gaactatatt gctttaaaat ctgaaggaga 2400
tgattactat attaatggtg cctggactat tgactggcct aggaaatttg atgttgctgg 2460
gacagctttt cattacaaga gaccaactga tgaaccagaa tccttggaag ctctaggtcc 2520
tacctcagaa aatctcatcg tcatggttct gcttcaagaa cagaatttgg gaattaggta 2580
taagttcaat gttcccatca ctcgaactgg cagtggagat aatgaagttg gctttacatg 2640
gaatcatcag ccttggtcag aatgctcagc tacttgtgct ggaggtgtcc aaagacagga 2700
ggtggtctgt aaaaggttgg atgacaactc cattgtccag aacaattact gtgatcctga 2760
cagtaagcca cctgaaaatc aaagagcctg caacactgag ccctgcccac ctgagtggtt 2820
cattggggat tggttggaat gcagcaagac ttgtgatggt gggatgcgca caagggcagt 2880
gctctgcatc aggaagatcg gaccttctga ggaggagacg ctggactaca gtggttgttt 2940
aacacaccgg cctgtcgaaa aagagccctg caacaaccag tcatgtccac cacagtgggt 3000
ggctttggac tggtctgagt gtactccaaa atgtggtcca ggattcaagc atcggattgt 3060
tctgtgcaag agcagtgacc tttctaagac attcccagct gcacaatgtc cagaggaaag 3120
caaacctcct gtccgcatcc gctgcagttt gggccgctgc cctcctcctc gctgggtcac 3180
aggagactgg ggccagtgtt ctgctcagtg tggccttgga cagcagatga gaactgtgca 3240

gtgtctctcc tacaccggac aggcatctag tgactgtcta gaaactgttc ggcctccatc 3300
aatgcagcag tgtgaaagca aatgtgacag tacccccatt tctaatactg aagagtgcaa 3360
agatgtgaat aaagtggctt attgcccact ggtgctgaag ttcaagttct gcagtcgagc 3420
atacttcaga cagatgtgtt gtaagacctg ccaaggacac tgacccacag aaagccagag 3480
agagtgcctt gtcatttcat catggaaatg catccatcaa agagagccac ccagaggaag 3540
aggattgatg tccttgcaaa tgcattaccc tgtggaaaac gtaaccactg gtcagcccta 3600
gctgacaaaa tttcaatatt attttagctt ctgtgaagtg ggatttattg atccaaagtg 3660
ctggacacgg tattaggagg gaatgccaga ttggagagat ccaaacaaca cagggagact 3720
tgcttactgt ggagcgtttg tgttctttcg agtaaatcca atagcctgtt tacctccttg 3780
gaccattaag ataattttta ttatggactt agcaatgaca ctgaatccat ttgtatttaa 3840
aactgtttaa aatgtagctg ttatgacttg gtcaactatg gaagtgaaga aggttcagaa 3900
ttcttaagtc atagcttaaa aatatttact gtactttatc tcactacaac agcaccacaa 3960
tttaaattat aaaacgggct ttgaactata atttaaggag caattataaa tcaaaagtaa 4020
tgaaagtttg tattattttt cttcattcca cttaatttcc ttaggaataa tcccctggtt 4080
ctgaacactg ctgtgagcca tatataaaac tatattaaac tgaacaataa tgaggggcat 4140
agtttaaagc agtgcatcag ttactgcagc tgtgcaagtc tataaactca gtgctgaaag 4200
actgtggcca acttgccatt gtgcaagtaa agctgagatt tccattaaaa ctttaagaga 4260
aaaacatttc aatttcatgc agaaaccaga cctggggtat ggtacagacc aaaggaccag 4320
gccctttgct gccaccacac aggatgcctt agttcttatt tgagtccctc caactcactt 4380
gtgtttacat cctccccagc cacagcacgg cttctgccct ttggattgct gcacgtgtgt 4440
tgagcttact gagatgatac catgcaaaag atagactggc tcggtaacca ggcagaccct 4500
tttgcagttt gttgacaatt acgatgagtt ccagatgtcc cttctttgat atggtagaag 4560
ggcatttatt tatatgagag caaatgtgtg tgtgtgtttg cgggcgcttt taagtgtgtg 4620
gatagatgag tgtgcttgca cataatgtgc tatttctgtg agttttaaag taggcaaggg 4680
ataataacca aagaagaaaa tttcatgaag actagacatc ataaagcata attttaatag 4740
tcactcaacc aagtattttt tatttttttat ggatactctg aatggcaatt aaatgtgaaa 4800
cccagtttct tgggcaagtc aaattctgga atcacatcca cctaaattaa aatgactagc 4860
tcgtattttc cccatcttca agtttcacat cctggtcatc aaaagactcg acagcaagac 4920
ttagaatgaa aaagggtact tgtttatatt aatatttttt acttgaacac gtgtagcttg 4980
cagcaggttc ttgatgaatg tgctttgtgt ccaaaatgcc tccccattgt acacaggtgt 5040
acaccatgca tgcaccaaca cctaaaactc aaaactaaat ggctattttg taaggttaat 5100
actttcagtt aaacagcatg tttgacttga ttccatcatg gtgctcttaa attacatgtc 5160
agtgcatcac atatatcatg atctaatgca gatgactagg ctttttccaa aaggaagaca 5220
gaccctcaga caccaaaagc caatctaaac aactcccagg tttgctgtgg acaatcagca 5280
tggaatgttt tctgcactct cagtcatgac catctgtatc ttgttacctg ctttctctct 5340
caacaccaca gttctcaacc ctgagccttc cagagagagc tattgatgat acaagaggaa 5400
tcaccagggc ccggatctaa gatgccctta gaagaccagc ccaagtgccg tcttagccat 5460
tcagtgaagg gcaaacagcc catgggtagt atggcccgag cactgaattc ccttgcgcct 5520
tttcaaagaa cagttaactt ggtgctaatg tgccctggtg aaataaataa aagatgggca 5580
gtttctgtgg cattttaggc ataggtttgc aatccagatc tgatttttctc caacataaat 5640
atcagctcat gttcttattt caaaaagatt tcttattacc gactaaaagc tatttttttac 5700
ctcacctgga aactaccatt gtgagggcca tcccccaggc actgcacagc accttggctg 5760
atgctggaag aggagggcag tcagtgtcac ttctgggatg tgccccagca ctgagaacaa 5820
aatgcaggca tccccccgggg cagcatcaga gtgcctttct agagggagcc acgcacagaa 5880
tgtaacagga tgaaacagtt tcaagtaagc cttgaattga aacctgagta ggttaaaaca 5940
attctatttc atagcacatc acaatactgc tgctactctg tagccacccc catggctaca 6000
tgatgcccta ttcctaaata ataacaatag cattgtcagt ggaggctggg ccaccatggc 6060
agaccttcca aaagtagtga gctacataga ctacttaggg aaccccaggg aaactggtac 6120
cctacacctg ggagcagtat ctgccactgg gataaagtcc tactaaaaaa ggaacggtaa 6180
atgtacccta atgattaaac cccgtgagat acatatgatt tccaaatagt ccatttcatt 6240
aggaactttt ttgtttgaat gaatgtcaca taggtatcct cagtaacaca gaacgaaatt 6300
acctttgtat tattgtgatt agttgttgct tattatttta tactcagtaa taatgtggta 6360
cactgttaat ttttttgctt ttgtaaatta tattctaatt tattgccatg tttcctaaca 6420
cttgtcctac attcattctc ctgcttgtaa tgaaaatgaa aaaatcattg taacacttga 6480
tggagtgaaa ttccacgcca ggcacagaat ttttttgaca tagataattt agtaaaataa 6540
aaattcagct tataataatg aaaaa 6565

<210> 23
<211> 1130
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 7480224CB1

<400> 23
gcgggtgaag accaaaggag aggaggggt gaagcagagg aatccatcta ggagaagcta 60
gttctggcag ctccccattg gcctcttcct gggagcctga gtccgggaag caggaagcgc 120
tcactggctc tgaggacaga gacatgggcc ctgctggctg tgccttcacg ctgctccttc 180
tgctggggat ctcagtgtgt gggcagcctg tatactccag ccgcgttgtg ggtggccagg 240
atgctgctgc agggcgctgg ccttggcagg tcagcctaca ctttgaccac aactttatct 300
atggaggttc cctcgtcagt gagaggttga tactgacagc agcacactgc atacaaccga 360
cctggactac tttttcatat actgtgtggc taggatcgat tacagtaggt gactcaagga 420
aacgtgtgaa gtactacgtg tccaaaatcg tcatccatcc caagtaccaa gatacaacgg 480
cagacgtcgc cttgttgaaa ctgtcctctc aagtcacctt cacttctgcc atcctgccta 540
tttgcttgcc cagtgtcaca aagcagttgg caattccacc cttttgttgg gtgaccggat 600
ggggaaaagt taaggaaagt tcagatagag attaccattc tgcccttcag gaagcagaag 660
tacccattat tgaccgccag gcttgtgaac agctctacaa tcccatcggt atcttcttgc 720
cagcactgga gccagtcatc aaggaagaca agatttgtgc tggtgatact caaaacatga 780
aggatagttg caagggtgat tctggagggc ctctgtcgtg tcacattgat ggtgtatgga 840
tccagacagg agtagtaagc tggggattag aatgtggtaa atctcttcct ggagtctaca 900
ccaatgtaat ctactaccaa aaatggatta atgccactat ttcaagagcc aacaatctag 960
acttctctga cttcttgttc cctattgtcc tactctctct ggctctcctg cgtccctcct 1020
gtgcctttgg acctaacact atacacagag taggcactgt agctgaagct gttgcttgca 1080
tacagggctg ggaagagaat gcatggagat ttagtcccag gggcagataa 1130

<210> 24
<211> 2372
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481056CB1

<400> 24
tcctggtaat ggttcatgat gtacgcacct gttgaatttt cagaagctga attctcacga 60
gctgaatatc aaagaaagca gcaattttgg gactcagtac ggctagctct tttcacatta 120
gcaattgtag caatcatagg aattgcaatt ggtattgtta ctcattttgt tgttgaggat 180
gataagtctt tctattacct tgcctctttt aaagtcacaa atatcaaata taaagaaaat 240
tatggcataa gatcttcaag agagtttata gaaaggagtc atcagattga aagaatgatg 300
tctaggatat ttcgacattc ttctgtaggc ggtcgattta tcaaatctca tgttatcaaa 360
ttaagtccag atgaacaagg tgtggatatt cttatagtgc tcatatttcg atacccatct 420
actgatagtg ctgaacaaat caagaaaaaa attgaaaagg ctttatatca aagtttgaag 480
accaaacaat tgtctttgac cataaacaaa ccatcattta gactcacacg ctgtggaata 540
aggatgacat cttcaaacat gccattacca gcatcctctt ctactcaaag aattgtccaa 600
ggaagggaaa cagctatgga aggggaatgg ccatggcagg ccagcctcca gctcataggg 660
tcaggccatc agtgtggagc cagcctcatc agtaacacat ggctgctcac agcagctcac 720
tgcttttgga aaaataaaga cccaactcaa tggattgcta cttttggtgc aactataaca 780
ccacccgcag tgaaacgaaa tgtgaggaaa attattcttc atgagaatta ccatagagaa 840
acaaatgaaa atgacattgc tttggttcag ctctctactg gagttgagtt ttcaaatata 900
gtccagagag tttgcctccc agactcatct ataaagttgc cacctaaaac aagtgtgttc 960
gtcacaggat ttggatccat tgtagatgat ggacctatac aaaatacact tcggcaagcc 1020
agagtggaaa ccataagcac tgatgtgtgt aacagaaagg atgtgtatga tggcctgata 1080
actccaggaa tgttatgtgc tggattcatg gaaggaaaaa tagatgcatg taagggagat 1140
tctggtggac ctctggttta tgataatcat gacatctggt acattgtagg tatagtaagt 1200
tggggacaat cgtgtgcact tcccaaaaaa cctggagtct acaccagagt aactaagtat 1260
cgagattgga ttgcctcaaa gactggtatg tagtgtggat tgtccatgag ttatacacat 1320
ggcacacaga gctggtactc ctgcgtattt tgtattgttt aaattcattt actttggatt 1380
agtgcttttg ctagatgtca agaagccctt cagacccaga caaatctaat atcctgaggt 1440
ggcctttaca tacgtaggac caaacccccc ctaccatgag ggaagaagac acagcaaatg 1500
acagacagca cctattcctt actcacaagg gaaactgctt gtgatacttc ctaataagat 1560
aaataagtgg tttccctcaa ttgaagacag gaacatcatt ttccacagga tatgaagagc 1620
tgccagtaat gccaaaatct tacctcatat aatacctgga gcatgtgaga ttcttctagt 1680
gaaaaagaac agtcttccct gaagactcag ggcttcaaca ttctagaact gataagtgga 1740

ccttcagtgt gcaagaatgg agaagcatgg gatttgcatt atgacttgaa ctgggcttat 1800
atctaataat acagagcact atcactaacc tcaacagttg acattttaaa agtttttaaa 1860
tgtatctgaa cttgctgtta acacagtgtt ataactcaag cactagcttc aggaagcatg 1920
ttgtgttgtt aagaagcttt tctgatttat tctttaacag catcttgcca tctatatgtt 1980
agtagcagtt ggcccagaaa ggacgaaaaa aagattaaga ctctttggaa cgtttttcca 2040
tgagcacagg aggataaaaa gaagcagatg aaggctagga gaattggttt caaataatta 2100
gtaacaggac aagcacgcta atttttgatg gaatgagtta tccaattatt tacttagaaa 2160
tatttatatc agtatatggc aactggtact tttgtaagtc ttcagctttc tgacaagtca 2220
gatgtccatc agagtatcag gtcaggtgtc tatcagaata tcagagctga tttgtgtaaa 2280
gcttgtgtaa agcacgtagg acagtgcctt gcatatacta cgaactaaat aaatctttgt 2340
tatatggaaa tcaaaaaaaa aaaaaaaaaa aa 2372

<210> 25
<211> 4253
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3750264CB1

<400> 25
tgaggactga gggtcttagg gggaccggga cagacccaaa gacactctag acaagaccag 60
agagagcccc tgaaggagga ggatggggca ccaggcctgg caatgcaaga acaggagagg 120
agggagggag ccagtgggag aaaggggtga ggtccctgct tcacttgcaa tgagaatgtt 180
cctacctttc aggggtggct cagggcagga gcgggggtca gaggtgccca accaggaagg 240
gccttgatct gggagttggc tgacacttcc aaagaaggaa tagggaagaa gaagcaagaa 300
gagagggaga gggagaggag gtgggttttt tgttggaggg ggttcattag gaacagaaga 360
aagaagaagt ctaagaggaa gttctccagg ggcagagaga gggtcagaat ttcctcagtg 420
atccctcaac tacagaccca gctcagtgct gaagaccagc ccggctcctc ctctttgacc 480
cctccctgcc caggctccaa agaagaagaa accaaggccc agagagggag gcccaggtgc 540
agggagcagg cgagggaagg atccgtacag gggcccaaca ctactccacc aaccgaagcc 600
cccaaaagga gcccggtgat gctgcgaagg ctgtgaacag gggaggcggc actgtggggg 660
ctgccggcag ccggggctgg ggagagacat gtggacacgt ggcctctatg gctcccgcct 720
gccagatcct ccgctgggcc ctcgccctgg ggctgggcct catgttcgag gtcacgcacg 780
ccttccggtc tcaagatgag ttcctgtcca gtctggagag ctatgagatc gccttcccca 840
cccgcgtgga ccacaacggg gcactgctgg ccttctcgcc acctcctccc cggaggcagc 900
gccgcggcac gggggccaca gccgagtccc gcctcttcta caaagtggcc tcgcccagca 960
cccacttcct gctgaacctg acccgcagct cccgtctact ggcagggcac gtctccgtgg 1020
agtactggac acgggagggc ctggcctggc agagggcggc ccggccccac tgcctctacg 1080
ctggtcacct gcagggccag gccagcagct cccatgtggc catcagcacc tgtggaggcc 1140
tgcacggcct gatcgtggca gacgaggaag agtacctgat tgagcccctg cacggtgggc 1200
ccaagggttc tcggagcccg gaggaaagtg gaccacatgt ggtgtacaag cgttcctctc 1260
tgcgtcaccc ccacctggac acagcctgtg gagtgagaga tgagaaaccg tggaaagggc 1320
ggccatggtg gctgcggacc ttgaagccac cgcctgccag gcccctgggg aatgaaacag 1380
agcgtggcca gccaggcctg aagcgatcgg tcagccgaga gcgctacgtg gagaccctgg 1440
tggtggctga caagatgatg gtggcctatc acgggcgccg ggatgtggag cagtatgtcc 1500
tggccgtcat gaacattgtt gccaaacttt tccaggactc gagtctggga agcaccgtta 1560
acatcctcgt aactcgcctc atcctgctca cggaggacca gcccactctg gagatcaccc 1620
accatgccgg gaagtccctg gacagcttct gtaagtggca gaaatccatc gtgaaccaca 1680
gcggccatgg caatgccatt ccagagaacg gtgtggctaa ccatgacaca gcagtgctca 1740
tcacacgcta tgacatctgc atctacaaga acaaaccctg cggcacacta ggcctggccc 1800
cggtgggcgg aatgtgtgag cgcgagagaa gctgcagcgt caatgaggac attggcctgg 1860
ccacagcgtt caccattgcc cacgagatcg ggcacacatt cggcatgaac catgacggcg 1920
tgggaaacag ctgtggggcc cgtggtcagg acccagccaa gctcatggct gcccacatta 1980
ccatgaagac caacccattc gtgtggtcat cctgcagccg tgactacatc accagctttc 2040
tagactcggg cctggggctc tgcctgaaca accggccccc cagacaggac tttgtgtacc 2100
cgacagtggc accgggccaa gcctacgatg cagatgagca atgccgcttt cagcatggag 2160
tcaaatcgcg tcagtgtaaa tacggggagg tctgcagcga gctgtggtgt ctgagcaaga 2220
gcaaccggtg catcaccaac agcatcccgg ccgccgaggg cacgctgtgc cagacgcaca 2280
ccatcgacaa ggggtggtgc tacaaacggg tctgtgtccc ctttgggtcg cgcccagagg 2340
gtgtggacgg agcctggggg ccgtggactc catggggcga ctgcagccgg acctgtggcg 2400
gcggcgtgtc ctcttctagc cgtcactgcg acagccccag gccaaccatc gggggcaagt 2460

actgtctggg tgagagaagg cggcaccgct cctgcaacac ggatgactgt ccccctggct 2520
cccaggactt cagagaagtg cagtgttctg aatttgacag catccctttc cgtgggaaat 2580
tctacaagtg gaaaacgtac cggggagggg gcgtgaaggc ctgctcgctc acgtgcctag 2640
cggaaggctt caacttctac acggagaggg cggcagccgt ggtggacggg acaccctgcc 2700
gtccagacac ggtggacatt tgcgtcagtg gcgaatgcaa gcacgtgggc tgcgaccgag 2760
tcctgggctc cgacctgcgg gaggacaagt gccgagtgtg tggcggtgac ggcagtgcct 2820
gcgagaccat cgagggcgtc ttcagcccag cctcacctgg ggccgggtac gaggatgtcg 2880
tctggattcc caaaggctcc gtccacatct tcatccagga tctgaacctc tctctcagtc 2940
acttggccct gaagggagac caggagtccc tgctgctgga ggggctgccc gggacccccc 3000
agccccaccg tctgcctcta gctgggacca cctttcaact gcgacagggg ccagaccagg 3060
tccagagcct cgaagccctg ggaccgatta atgcatctct catcgtcatg gtgctggccc 3120
ggaccgagct gcctgccctc cgctaccgct tcaatgcccc catcgcccgt gactcgctgc 3180
ccccctactc ctggcactat gcgccctgga ccaagtgctc ggcccagtgt gcaggcggta 3240
gccaggtgca ggcggtggag tgccgcaacc agctggacag ctccgcggtc gccccccact 3300
actgcagtgc ccacagcaag ctgcccaaaa ggcagcgcgc ctgcaacacg gagccttgcc 3360
ctccagactg ggttgtaggg aactggtcgc tctgcagccg cagctgcgat gcaggcgtgc 3420
gcagccgctc ggtcgtgtgc cagcgccgcg tctctgccgc ggaggagaag gcgctggacg 3480
acagcgcatg cccgcagccg cgcccacctg tactggaggc ctgccacggc cccacttgcc 3540
ctccggagtg ggcggccctc gactggtctg agtgcacccc cagctgcggg ccgggcctcc 3600
gccaccgcgt ggtcctttgc aagagcgcag accaccgcgc cacgctgccc ccggcgcact 3660
gctcacccgc cgccaagcca ccggccacca tgcgctgcaa cttgcgccgc tgccccccgg 3720
cccgctgggt ggctggcgag tggggtgagt gctctgcaca gtgcggcgtc gggcagcggc 3780
agcgctcggt gcgctgcacc agccacacgg gccaggcgtc gcacgagtgc acggaggccc 3840
tgcggccgcc caccacgcag cagtgtgagg ccaagtgcga cagcccaacc cccggggacg 3900
gccctgaaga gtgcaaggat gtgaacaagg tcgcctactg ccccctggtg ctcaaatttc 3960
agttctgcag ccgagcctac ttccgccaga tgtgctgcaa aacctgccag ggccactagg 4020
gggcgcgcgg cacccggagc cacagctggc ggggtctccg ccgccagccc tgcagcgggc 4080
cggccagagg gggccccggg ggggcgggaa ctgggaggga agggtgagac ggagccggaa 4140
gttatttatt gggaacccct gcagggccct ggctgggggg atggagaggg gctggctatc 4200
cccccagagc ccctcttcag catccgcccc ttccagttca catagtgaga ccc 4253

<210> 26
<211> 2681
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1749735CB1

<400> 26
ggatattaat gaaaaaattt gaatcaatac acagaggcaa gaaaagaaaa aaagaattgt 60
gatccgtatg ctcacatgct tttccttgac ctaacatagc aaatacccca tccacctttt 120
tcctttccaa gagaccatat aaatgaacaa acaaaagctc tggcgaaaca agccagctgt 180
gtcccgcccc cttctggctt gctgctgggc tttgtgacac ttaacttaca ttctcaccaa 240
cttttcagca ggatgctcgc gaaaatcttg ttattagtgt ttaagaaagt aacctccttt 300
atatttttta caatagcatt ggtttttgtt tgatatgtta tagtttacag agggctttat 360
taaagtacat tatgatcatt ctctcttaac aaccatgcct tgagataggt agcttgtagt 420
ctccatttag agtttggaag ctacagcagc aaagtgacta ttgcacaccc aataaatggc 480
agagtcagga ttggattcta aatccagggt ctttctgctg catcagagct gccaccttct 540
cacccttaa aaacatgatg gtggccgggc acagtggctc acacctgtga tatcagcact 600
ttgggaggct gaggcaggag ttcaacacca gctggggcaa catagtgaga cctcatctct 660
acaaaacaaa aaacaagaaa acctgacgta aacataatgt ttttaacttt tgttgtgctg 720
acttctctca ctcccctatg gagtggaaat gcctgtgtga gatccataga tgcttttcct 780
cctcaacagt tccaccatgc catattcaca ttaggatatg attctcctgc taaatcatct 840
gtacatcaga tgtacacatc aattgtgggc cctaggtgct tatctgcaac acattgcttc 900
tctgtttttt tactgctcaa gtgctctgag atgaatcctt ctaattagcc tctctcctta 960
aaagttctaa gactctttct caaactagga tgtatgcact atttggacca gaatcaccca 1020
gagggcttat taaaaacgca tattccagga cccaccttac acttgataca gaatgtctgg 1080
gagtgggacc agggaatctg aatttttatt aggcttctca aataatttta agaattccaa 1140
ggtttgagaa atgatctaag atacctatgt gttgtgctgt aatttttgtg accttccctt 1200
gatttaattt acttttctac ttagtttact tgaagcctaa cccaatctca gcatctcttt 1260
tctaactcca agagccattg tttcattctt gaagaatgaa aaccttagag ttcccttaaa 1320

ctgctaagta aagatactgt ggaatttctg gtgctctgtc caaaatccag cgtctttgct 1380
gatgactagg taagaggaag cttaaggagc ctgccttaaa gcagaggaag atctgaaatc 1440
attgcactga agaagcaaga ctgactttgg tttgttttta agagagaggc ccaaggaatc 1500
cagctgcctc acactggggt ggagttgctg ggaagggtct gtagcaggca tgtgcttcat 1560
gctgtgggcc agagccatta gggagatctc ttcacagagc tgtcagggag atcagttcag 1620
aggccattcc cacctgaggt aacacagtgc cgacacctct tcctgggatt cctcaaaagt 1680
gtcacctcac ctggacagtt ttattctttt ctaggtaatt agaactcagt attctagaat 1740
gtggaggctt agcacccaaa atttaggtga agggttgatg agtttgggct ttaacattta 1800
ccttgtgaca ggatgaagca cttcaacttg ccaagtcttg tttttctcat ctgtaaaata 1860
ataatactaa tatctgccct gtctgctata ctgccgtttt tgtgaagatg aagtgagaag 1920
gatatatgag aacaaggtgg cagttatcga gagagaactc aaggtctcca gcatgcaggt 1980
tttcactgag cagcttctga aacccttaca aagcagccag cggcttttgt gcagaggagt 2040
gccacttcct tcagagagag aacacggttt tcctttcttc ctctttccct cttccgttca 2100
actcttgtag aagccaaaca ccagatacat aatgtcctaa tgcccctgct tccggacctg 2160
ttttcgttgt tggggttttt cctccctgct gggtcctcca gctgggtcac agtgtgctcg 2220
tgttcttcct gcctctgagg ccacttccct ggttggcgtg tctcctgtgg ccgcacgcct 2280
tctgtgttat ccctgatagc tgtgttgtgg acttcccagc atgcgccatc cgtgaacgtg 2340
gtatcatggt gaggcagaaa ggcagcttct tacccccatc attcagatga ggagatgaga 2400
tgctgtgtca ggggcacatc atttcttcct tgggccctgt gcttggaccc aagctgtgcc 2460
gtcctgtcat ctagcccccg tgccctttcc accagtgaca cctgcagctc agttagcacg 2520
aggcccttga gttatattca gtatcctttg tccccactat aaagctgaat gtctaaaatc 2580
ctccccccta ctcccctttgg ttactttcta ttttaaatat tcttgtaggt ggatttacat 2640
caccttcatt ttaaaataac ccctctctta aaggtaaaaa a 2681

<210> 27
<211> 4506
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473634CB1

<400> 27
atggtgacca tctgcctggt cactgcctgg acaggactct cctggtctta tcacctaaga 60
tcccatatcc tggaaacccc cctgatagta gaaaaccgga atatttggac ctctaatgaa 120
cgggacagag gctcccaaag tgttgggact acaggcatca gccaccgcgc caagcctgta 180
tcttgtttct taaaatacaa agcaactgag ggagcctgcg gaggaacctt acgcgggacc 240
agcagctcca tctccagccc gcacttccct tcagagtacg agaacaacgc ggactgcacc 300
tggaccattc tggctgagcc cggggacacc attgcgctgg tcttcactga ctttcagcta 360
gaagaaggat atgatttctt agagatcagt ggcacggaag ctccatccat atggctaact 420
ggcatgaacc tcccctctcc agttatcagt agcaagaatt ggctacgact ccatttcacc 480
tctgacagca accaccgacg caaaggattt aacgctcagt tccaagtgaa aaaggcgatt 540
gagttgaagt caagaggagt caagatgctg cccagcaagg atggaagcca taaaaactct 600
gtcttgagcc aaggaggtgt tgcattggtc tctgacatgt gtccagatcc tgggattcca 660
gaaaatggta gaagagcagg ttccgacttc agggttggtg caaatgtaca gttttcatgt 720
gaggacaatt acgtgctcca gggatctaaa agcatcacct gtcagagagt tacagagacg 780
ctcgctgctt ggagtgacca caggcccatc tgccgagcga gaacatgtgg atccaatctg 840
cgtgggccca gcggcgtcat tacctcccct aattatccgg ttcagtatga agataatgca 900
cactgtgtgt gggtcatcac caccaccgac ccggacaagg tcatcaagct tgcctttgaa 960
gagtttgagc tggagcgagg ctatgacacc ctgacggttg gtgatgctgg gaaggtggga 1020
gacaccagat cggtcttgta cgtgctcacg ggatccagtg ttcctgacct cattgtgagc 1080
atgagcaacc agatgtggct acatctgcag tcggatgata gcattggctc acctgggttt 1140
aaagctgttt accaagaaat tgaaaaggga gggtgtgggg atcctggaat ccccgcctat 1200
gggaagcgga cgggcagcag tttcctccat ggagatacac tcacctttga atgcccggcg 1260
gcctttgagc tggtggggga gagagttatc acctgtcagc agaacaatca gtggtctggc 1320
aacaagccca gctgtgtatt ttcatgtttc ttcaacttta cggcatcatc tgggattatt 1380
ctgtcaccaa attatccaga ggaatatggg aacaacatga actgtgtctg gttgattatc 1440
tcggagccag gaagtcgaat tcacctaatc tttaatgatt ttgatgttga gcctcaattt 1500
gactttctcg cggtcaagga tgatggcatt tctgacataa ctgtcctggg tacttttttct 1560
ggcaatgaag tgccttccca gctggccagc agtgggcata tagttcgctt ggaatttcag 1620
tctgaccatt ccactactgg cagagggttc aacatcactt acaccacatt tggtcagaat 1680
gagtgccatg atcctggcat tcctataaac ggacgacgtt ttggtgacag gtttctactc 1740

gggagctcgg tttctttcca ctgtgatgat ggctttgtca agacccaggg atccgagtcc 1800
attacctgca tactgcaaga cgggaacgtg gtctggagct ccaccgtgcc ccgctgtgaa 1860
gctccatgtg gtggacatct gacagcgtcc agcggagtca ttttgcctcc tggatggcca 1920
ggatattata aggattcttt acattgtgaa tggataattg aagcaaaacc aggccactct 1980
atcaaaataa cttttgacag atttcagaca gaggtcaatt atgacacctt ggaggtcaga 2040
gatgggccag ccagttcgtc cccactgatc ggcgagtacc acggcaccca ggcaccccag 2100
ttcctcatca gcaccgggaa cttcatgtac ctgctgttca ccactgacaa cagccgctcc 2160
agcatcggct tcctcatcca ctatgagagt gtgacgcttg agtcggattc ctgcctggac 2220
ccgggcatcc ctgtgaacgg ccatcgccac ggtggagact ttggcatcag gtccacagtg 2280
actttcagct gtgacccggg gtacacacta agtgacgacg agcccctcgt ctgtgagagg 2340
aaccaccagt ggaaccacgc cttgcccagc tgcgacgctc tatgtggagg ctacatccaa 2400
gggaagagtg gaacagtcct ttctcctggg tttccagatt tttatccaaa ctctctaaac 2460
tgcacgtgga ccattgaagt gtctcatggg aaaggagttc aaatgatctt tcacaccttt 2520
catcttgaga gttcccacga ctatttactg atcacagagg atggaagttt ttccgagccc 2580
gttgccaggc tcaccgggtc ggtgttgcct catacgatca aggcaggcct gtttggaaac 2640
ttcactgccc agcttcggtt tatatcagac ttctcaattt cgtacgaggg cttcaatatc 2700
acattttcag aatatgacct ggagccatgt gatgatcctg gagtccctgc cttcagccga 2760
agaattggtt ttcactttgg tgtgggagac tctctgacgt tttcctgctt cctgggatat 2820
cgtttagaag gtgccaccaa gcttacctgc ctgggtgggg gccgccgtgt gtggagtgca 2880
cctctgccaa ggtgtgtggc cgaatgtgga gcaagtgtca aaggaaatga aggaacatta 2940
ctgtctccaa attttccatc caattatgat aataaccatg agtgtatcta taaaatagaa 3000
acagaagccg gcaagggcat ccaccttaga acacgaagct tccagctgtt tgaaggagat 3060
actctaaagg tatatgatgg aaaagacagt tcctcacgtc cactgggcac gttcactaaa 3120
aatgaacttc tggggctgat cctaaacagc acatccaatc acctgtggct agagttcaac 3180
accaatggat ctgacaccga ccaaggtttt caactcacct ataccagttt tgatctggta 3240
aaatgtgagg atccgggcat ccctaactac ggctatagga tccgtgatga aggccacttt 3300
accgacactg tagttctgta cagttgcaac ccggggtacg ccatgcatgg cagcaacacc 3360
ctgacctgtt tgagtggaga caggagagtg tgggacaaac cactaccttc gtgcatagcg 3420
gaatgtggtg gtcagatcca tgcagccaca tcaggacgaa tattgtcccc tggctatcca 3480
gctccgtatg acaacaacct ccactgcacc tggattatag aggcagaccc aggaaagacc 3540
attagcctcc atttcattgt tttcgacacg gagatggctc acgacatcct caaggtctgg 3600
gacgggccgg tggacagtga catcctgctg aaggagtgga gtggctccgc ccttccggag 3660
gacatccaca gcaccttcaa ctcactcacc ctgcagttcg acagcgactt cttcatcagc 3720
aagtctggct tctccatcca gttctccaga tctcaggctg gaacacgaag acgctggtct 3780
gaccacccca aagccagtca ttcagctact ctccacaaaa tgtagcttgc cacttctggg 3840
aaccagtgag aatcgggcac cagtctccat ctccctgaga acctgataaa catttgactc 3900
ctacacctgg aataaatcat gtcctggttt tctagtttta gaaaagaagg ttcctataac 3960
ccctcagtcg taattaagaa actgacccag ttaccctgct tcactgcagg aagaaactgg 4020
gctgttatgt ccctctcact ccacccacat tcgtcccctc actggcgaat ccagccatga 4080
aactaaatca agctggtgtc ttcccaaacc aaaggtggga aactcttcac aaagtgcaaa 4140
acagcctgtc catcacacca agaagccatc actactcttt tgtaggtggg aggatggggt 4200
gggacgatgg acatctctca ttttttgtct ttaatgaacc tgcgaccaca aaaaatgagg 4260
acttacctat atacgatggt gtgtgctcca ttaccctgct aatttttact tcaaacgtgg 4320
cattgttctg atttcacatg ttaactgacc caagaacgtt ccccccttatg aggttaaggg 4380
cccggttccc gcacaggcct tccgtttaag agacgcggca tcgccttcca cggaacactg 4440
ggctttgtga aacaaaaggg cgggccgcaa ccgcgggaat acaccgccac acgacacggc 4500
gacacc 4506

<210> 28
<211> 1125
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4767844CB1

<400> 28
ggaattccag agctgccagg cgctcccagc cggtctcggc aaacttttcc ccagcccacg 60
tgctaaccaa gcggctcgct tcccgagccc gggatggagc accgcgccta gggaggccgc 120
gccgcccgag acgtgcgcac ggttcgtggc ggagagatgc tgatcgcgct gaactgaccg 180
gtgcggcccg ggggtgagtg gcgagtctcc ctctgagtcc tccccagcag cgcggccggc 240
gccggctctt tgggcgaacc ctccagttcc tagactttga gaggcgtctc tcccccgccc 300

gaccgcccag atgcagtttc gccttttctc ctttgccctc atcattctga actgcatgga 360
ttacagccac tgccaaggca accgatggag acgcagtaag cgagctagtt atgtatcaaa 420
tcccatttgc aagggttgtt tgtcttgttc aaaggacaat gggtgtagcc gatgtcaaca 480
gaagttgttc ttcttccttc gaagagaagg gatgcgccag tatggagagt gcctgcattc 540
ctgcccatcc gggtactatg gacaccgagc cccagatatg aacagatgtg caagatgcag 600
aatagaaaac tgtgattctt gctttagcaa agacttttgt accaagtgca aagtaggctt 660
ttatttgcat agaggccgtt gctttgatga atgtccagat ggttttgcac cattagaaga 720
aaccatggaa tgtgtggaag gatgtgaagt tggtcattgg agcgaatggg gaacttgtag 780
cagaaataat cgcacatgtg gatttaaatg gggtctggaa accagaacac ggcaaattgt 840
taaaaagcca gtgaaagaca caataccgtg tccaaccatt gctgaatcca ggagatgcaa 900
gatgacaatg aggcattgtc caggagggaa gagaacacca aaggcgaagg agaagaggaa 960
caagaaaaag aaaaggaagc tgatagaaag ggcccaggag caacacagcg tcttcctagc 1020
tacagacaga gctaaccaat aaaacaagag atccggtaga tttttagggg tttttgtttt 1080
tgcaaatgtg cacaaagcta ctctccactc ctgcacactg gtgtg 1125

<210> 29
<211> 3062
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7487584CB1

<400> 29
aatgtgagag gggctgatgg aagctgatag gcaggactgg agtgttagca ccagtactgg 60
atgtgacagc aggcagagga gcacttagca gcttattcag tgtccgattc tgattccggc 120
aaggatccaa gcatggaatg ctgccgtcgg gcaactcctg gcacactgct cctctttctg 180
gctttcctgc tcctgagttc caggaccgca cgctccgagg aggaccggga cggcctatgg 240
gatgcctggg gcccatggag tgaatgctca cgcacctgcg ggggaggggc ctcctactct 300
ctgaggcgct gcctgagcag caagagctgt gaaggaagaa atatccgata cagaacatgc 360
agtaatgtgg actgcccacc agaagcaggt gatttccgag ctcagcaatg ctcagctcat 420
aatgatgtca agcaccatgg ccagttttat gaatggcttc ctgtgtctaa tgaccctgac 480
aacccatgtt cactcaagtg ccaagccaaa ggaacaaccc tggttgttga actagcacct 540
aaggtcttag atggtacgcg ttgctataca gaatctttgg atatgtgcat cagtggttta 600
tgccaaattg ttggctgcga tcaccagctg ggaagcaccg tcaaggaaga taactgtggg 660
gtctgcaacg gagatgggtc cacctgccgg ctggtccgag ggcagtataa atcccagctc 720
tccgcaacca aatcggatga tactgtggtt gcaattccct atggaagtag acatattcgc 780
cttgtcttaa aaggtcctga tcacttatat ctggaaacca aaaccctcca ggggactaaa 840
ggtgaaaaca gtctcagctc cacaggaact ttccttgtgg acaattctag tgtggacttc 900
cagaaatttc cagacaaaga gatactgaga atggctggac cactcacagc agatttcatt 960
gtcaagattc gtaactcggg ctccgctgac agtacagtcc agttcatctt ctatcaaccc 1020
atcatccacc gatggaggga gacggatttc tttccttgct cagcaacctg tggaggaggt 1080
tatcagctga catcggctga gtgctacgat ctgaggagca accgtgtggt tgctgaccaa 1140
tactgtcact attacccaga gaacatcaaa cccaaaccca agcttcagga gtgcaacttg 1200
gatccttgtc cagccagtga cggatacaag cagatcatgc cttatgacct ctaccatccc 1260
cttcctcggt gggaggccac cccatggacc gcgtgctcct cctcgtgtgg gggggacatc 1320
cagagccggg cagtttcctg tgtggaggag gacatccagg ggcatgtcac ttcagtggaa 1380
gagtggaaat gcatgtacac ccctaagatg cccatcgcgc agccctgcaa cattttgac 1440
tgccctaaat ggctggcaca ggagtggtct ccgtgcacag tgacgtgtgg ccagggcctc 1500
agataccgtg tggtcctctg catcgaccat cgaggaatgc acacaggagg ctgtagccca 1560
aaaacaaagc cccacataaa agaggaatgc atcgtaccca ctccctgcta taaacccaaa 1620
gagaaacttc cagtcgaggc caagttgcca tggttcaaac aagctcaaga gctagaagaa 1680
ggagctgctg tgtcagagga gccctcgttc atcccagagg cctggtcggc ctgcacagtc 1740
acctgtggtg tggggaccca ggtgcgaata gtcaggtgcc aggtgctcct gtctttctct 1800
cagtccgtgg ctgacctgcc tattgacgag tgtgaagggc ccaagccagc atcccagcgt 1860
gcctgttatg caggcccatg cagcggggaa attcctgagt tcaacccaga cgagacagat 1920
gggctctttg gtggcctgca ggatttcgac gagctgtatg actgggagta tgaggggttc 1980
accaagtgct ccgagtcctg tggaggaggt gtccaggagg ctgtggtgag ctgcttgaac 2040
aaacagactc gggagcctgc tgaggagaac ctgtgcgtga ccagccgccg gcccccacag 2100
ctcctgaagt cctgcaattt ggatccctgc ccagcaagtc ctgtcatcta ggaagaagca 2160
gtatcgactc agcatggaac gcctgcaacg ttctttgtta ggcaaccaag aggcctggct 2220
tctcatcctg ctgtcaccaa ctagctctgt ggcctagggc gaggtgtctg ccctttatgt 2280

ttccacatct gcaaagtgaa ctggttgtac ctgatgatct gagatcccat gacttgctca 2340
catgtcccat gattctttat tttgtaggca gaagcattaa acagctactc ctgctgctgt 2400
gtgctaatca ttcctgtaat ttctgttctg cttatttgcc attatttgaa aaacatgcaa 2460
aagggtcttt ctaaccacat tcctgtgttg taacaacacc caaatgctga ggcagtgccg 2520
aggagtcagt gcctgggact tgcttaaaac tgctgggact cgtggtccct aaacccttct 2580
ttgagcacca aaacgaatag gacatgagat gttacttctc attctcaaag tactaactat 2640
gtttaagtta caaaaggtta ggttatcctg tgacccttt gttgactcac agacaagaac 2700
agttgttgag cttaatgttg tcgcatttgc tccagataaa ctcaattctc tgatttccca 2760
ccagccaact gtcaagccaa caggcaagac ctctcactgg gcacagccag gagtttcttg 2820
ggtcgaccat acacattgaa acatttgtag aaggttgcta attgcaacaa taaaggggac 2880
caaagtataa tggcctaatc tcatccaaga gtcaaaacag attttccccc taaaaatgat 2940
aattgtatag aggtgccttt cctgtggaat atctcactct gatgtcagag aaaaatctct 3000
ccttcccttc tcctggtgtt caatgtatac agaaaataaa atgtgtttgg taggaaaaaa 3060
aa 3062

<210> 30
<211> 1908
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1468733CB1

<400> 30
tcggccgaga atgctttagt atattgaaat ctttaagagc agtagagctg aagttagaac 60
tcattatgat ccaccacgaa agcttatggc catgcagcgg ccaggtcctt atgacagacc 120
tggggctggt agagggtata acagcattgg cagaggagct ggctttgaga ggatgaggcg 180
tggtgcttat ggtggaggct atggaggcta tgatgattac aatggctata atgatggcta 240
tggatttggg tcagatagat ttggaagaga cctcaattac tgtttttcag gaatgtctga 300
tcacatacgg ggatggtggc tctactttcc agagcacaac aggacactgt gtacacatgc 360
ggggattacc ttacagagct actgagaatg acatttataa tttttttca ccgctcaacc 420
ctgtgagagt acacattgaa attggtcctg atggcagagt aactggtgaa gcagatgtcg 480
agttcgcaac tcatgaagat gctgtggcag ctatgtcaaa agacaaagca aatatgcaac 540
acagatatgt agaactcttc ttgaattcta cagcaggagc aagcggtggt gcttacgaac 600
acagatatgt agaactcttc ttgaattcta cagcaggagc aagcggtggt gcttatggta 660
gccaaatgat gggaggcatg ggcttgtcaa accagtccag ctacgggggc ccagccagcc 720
agcagctgag tgggggttac ggaggcggcg gcggcggggg aggcgggggc ctgggtgggg 780
gcctgggaaa tgtgcttgga ggcctgatca gcggggccgg gggcggcggc ggcggcggcg 840
gcggcggcgg cggtggtgga ggcggcggtg gcggtggaac ggccatgcgc atcctaggcg 900
gagtcatcag cgccatcagc gaggcggctg cgcagtacaa cccggagccc ccgcccccac 960
gcacacatta ctccaacatt gaggccaacg agagtgagga ggtccggcag ttccggagac 1020
tctttgccca gctggctgga gatgacatgg aggtcagcgc cacagaactc atgaacattc 1080
tcaataaggt tgtgacacga caccctgatc tgaagactga tggttttggc attgacacat 1140
gtcgcagcat ggtggccgtg atggatagcg acaccacagg caagctgggc tttgaggaat 1200
tcaagtactt gtggaacaac atcaaaaggt ggcaggccat atacaaacag ttcgacactg 1260
accgatcagg gaccatttgc agtagtgaac tcccaggtgc ctttgaggca gcagggttcc 1320
acctgaatga gcatctctat aacatgatca tccgacgcta ctcagatgaa agtgggaaca 1380
tggattttga caacttcatc agctgcttgg tcaggctgga cgccatgttc cgtgccttca 1440
aatctcttga caaagatggc actggacaaa tccaggtgaa catccaggag tggctgcagc 1500
tgactatgta ttcctgaact ggagccccag acccgccccc tcaccgcctt gctataggag 1560
tcacctggag cctcggtctc tcccagggcc gatcctgtct gcagtcacat ctttgtgggg 1620
cctgctgacc cacaagcttt tgttctctca gtacttgtta cccagcttct caacatccag 1680
ggcccaattt gccctgcctg gagttccccc tggctctagg acactctaac aagctctgtc 1740
cacgggtctc cccattccca ccaggccctg cacacaccca ctccgtaact ctcccctgta 1800
cctgtgccaa gcctagcact tgtgatgcct ccatgcccgg agggcctctc tcagttctgg 1860
gaggatgact ccagtcctga cgcctgggac accttcacgg gttggtac 1908

<210> 31
<211> 1917
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1652084CB1

<220>
<221> unsure
<222> 1864, 1882, 1897-1898, 1902
<223> a, t, c, g, or other

<400> 31
atgctacaga aaggtgaatg tggagtaagt gggctaactg gccctagtga acaagggtgt 60
atagaaaaac ccttgaaact agctacctca cggacacaaa atagcagctg cagtagtaga 120
cacatgcaga taacccaagt gttagaggaa gaagagggct ggtttcctct tgtggatctc 180
ttcttattag aagccttttc tagaagcctt ccagcaacct ctcctgtctt tctcgcagtc 240
ggcataaaaa tgggttctct cagcacagct aacgttgaat tttgccttga tgtgttcaaa 300
gagctgaaca gtaacaacat aggagataac atcttctttt cttcgctgag tctgctttat 360
gctctaagca tggtcctcct tggtgccagg ggagagactg aagagcaatt ggagaaggta 420
tggaattcct cagaggtgct tcattttagt catactgtag actcattaaa accagggttc 480
aaggactcac ctaagccaga ctctaactgt accctcagca ttgccaacag gctctacggg 540
acaaagacga tggcatttca tcagcaatat ttaagctgtt ctgagaaatg gtatcaagcc 600
aggttgcaaa ctgtggattt tgaacagtct acagaagaaa cgaggaaaac gattaatgct 660
tgggttgaaa ataaaactaa tggaaaagtc gcaaatctct ttggaaagag cacaattgac 720
ccttcatctg taatggtcct ggtgaatgcc atatatttca aaggacaatg gcaaaataaa 780
tttcaagtaa gagagacagt taaaagtcct tttcagctaa gtgagggtaa aaatgtaact 840
gtggaaatga tgtatcaaat tggaacattt aaactggcct ttgtaaagga gccgcagatg 900
caagttcttg agctgcccta cgttaacaac aaattaagca tgattattct gcttccagta 960
ggcatagcta atctgaaaca gatagaaaag cagctgaatt cggggacgtt tcatgagtgg 1020
acaagctctt ctaacatgat ggaaagagaa gttgaagtac acctccccag attcaaactt 1080
gaaattaagt atgagctaaa ttccctgtta aaacctctag gggtgacaga tctcttcaac 1140
caggtcaaag ctgatctttc tggaatgtca ccaaccaagg gcctatattt atcaaaagcc 1200
atccacaagt catacctgga tgtcagcgaa gagggcacgg aggcagcagc agccactggg 1260
gacagcatcg ctgtaaaaag cctaccaatg agagctcagt tcaaggcgaa ccacccccttc 1320
ctgttcttta taaggcacac tcataccaac acgatcctat tctgtggcaa gcttgcctct 1380
ccctaatcag atggggttga gtaaggctca gagttgcaga tgaggtgcag agacaatcct 1440
gtgactttcc cacggccaaa aagctgttca cacctcacac acctctgtgc ctcagtttgc 1500
tcatctgcaa aataggtcta ggatttcttc caaccatttc atgagttgtg aagctaaggc 1560
tttgttaatc atggaaaaag gtagacttat gcagaaagcc tttctggctt tcttatctgt 1620
ggtgtctcat ttgagtgctg tccagtgaca tgatcaagtc aatgagtaaa attttaaggg 1680
attagatttt cttgacttgt atgtatctgt gagatcttga ataagtgacc tgacatctct 1740
gcttaaagaa aaccagctga agggcttcaa ctttgcttgg atttttaaat attttccttg 1800
catatgtaaa tagaatgtgg tgagttttag ttcaaaattc tctcgagaga ataatacatg 1860
cggnattttt cgtttcgggg tngtgtgtgc tgtggtnngg tncttatctt tctgatg 1917

<210> 32
<211> 1936
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3456896CB1

<400> 32
atggcgccgc cagccgcccg cctcgccctg ctctccgccg cggcgctcac gctggcggcc 60
cggcccgcgc ctagccccgg cctcggcccc ggacccgagt gtttcacagc caatggtgcg 120
gattataggg gaacacagaa ctggacagca ctacaaggcg ggaagccatg tctgttttgg 180
aacgagactt tccagcatcc atacaacact ctgaaatacc ccaacgggga gggggggcctg 240
ggtgagcaca actattgcag aaatccagat ggagacgtga gcccctggtg ctatgtggca 300
gagcacgagg atggtgtcta ctggaagtac tgtgagatac ctgcttgcca gatgcctgga 360
aaccttggct gctacaagga tcatggaaac ccacctcctc taactggcac cagtaaaacg 420
tccaacaaac tcaccataca aacttgcatc agtttttgtc ggagtcagag gttcaagttt 480
gctgggatgg agtcaggcta tgcttgcttc tgtggaaaca atcctgatta ctggaagtac 540
ggggaggcag ccagtaccga atgcaacagc gtctgcttcg gggatcacac ccaaccctgt 600

ggtggcgatg gcaggatcat cctctttgat actctcgtgg gcgcctgcgg tgggaactac 660
tcagccatgt cttctgtggt ctattcccct gacttccccg acacctatgc cacggggagg 720
gtctgctact ggaccatccg ggttccgggg gcctcccaca tccacttcag cttcccccta 780
tttgacatca gggactcggc ggacatggtg gagcttctgg atggctacac ccaccgtgtc 840
ctagcccgct tccacgggag gagccgccca cctctgtcct tcaacgtctc tctggacttc 900
gtcatcttgt atttcttctc tgatcgcatc aatcaggccc agggatttgc tgttttatac 960
caagccgtca aggaagaact gccacaggag aggcccgctg tcaaccagac ggtggccgag 1020
gtgatcacgg agcaggccaa cctcagtgtc agcgctgccc ggtcctccaa agtcctctat 1080
gtcatcacca ccagccccag ccacccacct cagactgtcc caggatggac agtctatggt 1140
ctggcaactc tcctcatcct cacagtcaca gccattgtag caaagatact tctgcacgtc 1200
acattcaaat cccatcgtgt tcctgcttca ggggacctta gggattgtca tcaaccaggg 1260
acttcggggg aaatctggag catttttac aagccttcca cttcaatttc catctttaag 1320
aagaaactca agggtcagag tcaacaagat gaccgcaatc cccttgtgag tgactaaaaa 1380
ccccactgtg cctaggactt gaggtccctc tttgagctca aggctgccgt ggtcaacctc 1440
tcctgtggtt cttctctgac agactcttcc cctcctctcc ctctgcctcg gcctcttcgg 1500
ggaaaaccct cctcctacag actaggaaga ggcaccctgc tgccagggca ggcagagcct 1560
ggattcctcc tgcttcatcg attgcactta ggagagagac tcaaagccct ggggcccggc 1620
cctctctgca tctctctctg atctagctag cagtgggggt gtcaggacag tgaggctgag 1680
atgacagagg tggtcatggc tggcacaggg ctcaggtaca ttctagatgg ctgtcaggtg 1740
gtgggtagct ttagttacat tgaatttttc ttgcttctct atttttgtcc acacacaaat 1800
cagtttctcc tgatctttat gtcttggaac agggccagac agggagaact ctcaggtact 1860
cttgggagtt ggtcccatac aagtgcggac tcctggacat tagcgaggtg taaagagggc 1920
agtgtctgtg ctgccc 1936